

Study Title

FIFRA Environmental Model Validation Task Force

Final Report

Data Requirement

none

Authors

Russell L. Jones and Mark H. Russell, Editors

Study Completed On

April 27, 2001

Performing Laboratory (Organization)

The FIFRA Environmental Model Validation Task Force

c/o American Crop Protection Association
1156 Fifteenth Street, NW, Suite 400
Washington, DC 20005

Laboratory Project ID

none

STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS

No claim of confidentiality is made for any information contained in this study on the basis of its falling within the scope of FIFRA § 10(d)(1)(A), (B), or (C).

COMPANY: FIFRA Environmental Model Validation Task Force

COMPANY AGENT: Russell L. Jones

AGENT TITLE: Final Report Editor

SIGNATURE: _____

DATE: _____

The above statement supersedes all other statements of confidentiality that may occur elsewhere in this report.

GOOD LABORATORY PRACTICE COMPLIANCE STATEMENT

The undersigned hereby declare that the work to which this report refers was performed according to the procedures herein described and this report provides an accurate record of the results obtained. Good Laboratory Practice does not apply to this study.

Submitter/Sponsor: _____

Date: _____

Acknowledgements

The work of the FIFRA Environmental Model Validation Task Force was a collaborative project involving scientists from the crop protection industry and the U.S. Environmental Protection Agency. Funding for the task force came from Aventis CropScience, BASF, Bayer, Dow AgroSciences, Dupont, FMC, ISK Biosciences, Monsanto, Rohm and Haas, Syngenta, Uniroyal, and Valent.

Industry scientists on the task force included Richard Allen, K. Balu, John Carbone, Kathy Carr, Sunmao Chen, Wenlin Chen, Pete Coody, Dave Gustafson, Amy Hackett, William Harned, Pat Havens, Paul Hendley, Jill Holihan, Scott Jackson, Russell Jones, David Judge, Dennis Laskowski, Jim Lin, Jerry Lucietta, Gary Mangels, Elizabeth Owens, Allan Rose, Mark Russell, Piyush Singh, Ian van Wesenbeeck, and Jeff Wolt.

EPA scientists acting as advisors to the task force included Michael Barrett, R. David Jones, Henry Nelson, and Ron Parker.

Outside scientists working at contract firms employed by the task force making a significant technical contribution included Allan Moose, William Warren-Hicks, Amy Ritter, Piyush Singh, and Christienne Zdinak. Assisting in coordination were Rob Paulsen and Marty Williams.

The chair of the FIFRA Environmental Model Validation Task Force was Dennis Laskowski who was assisted administratively by K. Balu, Sunmao Chen, Gary Mangels, Amy Hackett, and Paul Hendley. Steve Ragone served as task force coordinator. The work of the contractors in preparing the data sets and conducting the simulations was overseen by Mark Russell and Jeff Wolt. The data base subcommittee was chaired by Gary Mangels who was assisted by Russell Jones, Scott Jackson, and Ron Parker. The input parameter subcommittee was chaired by Scott Jackson and members were Jim Lin, Russell Jones, Mark Russell, and Ron Parker. The statistics subcommittee was chaired by John Carbone. Other industry task force members participating on this subcommittee included Wenlin Chen. EPA participants included R. Dave Jones, Michael Barrett, and Sandy Bird. William Warren-Hicks of the Cadmus Group served as consultant to the statistics subcommittee. The advice of several other non-task force scientists are also acknowledged and include Dennis Helsel (USGS), Scott Ray and Peter Shrearer (Dow AgroSciences), Jeff Weathington (DuPont), and John Weeks (S.C. Johnson). Principal participants in the Plackett-Burman tool were Jeff Wolt, Steve Cryer, and Mike Fouch) and for the Monte-Carlo tool were Pat Havens and Mike Fouch.

The FIFRA Environmental Model Validation Task Force received helpful comments from a number of external scientists. These included Jerry Hatfield (who coordinated comments from external scientists), Dave Eckhardt, Art Hornsby, John Hutson, Suresh Rao, Ali M. Sadeghi, Charlie Smith, Soroosh Sorooshian, Don Wauchope, Jeff Wagenet, and Andy Ward.

Communications between task force members and with outside scientists was facilitated by a web site. Paul Hendley created and maintained the web site. The web site was especially important due to the large number of large electronic files generated as part of the work of the task force. Early drafts of this report have been placed on the web site and the final report will also be placed here.

Many people contributed to the writing of the final report. The primary contributors to Section 3.1 and the associated Appendix 1 were Gary Mangels and Russell Jones. Sections 3.2, 5.4, and Appendix 4 were written by Jeff Wolt, Steve Cryer, and Scott Jackson. Sections 3.3, 5.3, 5.5, and Appendices 7 and 8 were by John Carbon, William Warren-Hicks, and Pat Havens. Gary Mangels and Jeff Wolt produced Chapter 4. Section 5.2 and Appendix 3 were written by Mark Russell and Jeff Wolt. The primary authors of Chapter 6 and Chapter 7 were Piyush Singh and Mark Russell, respectively. Appendix 2 was written by Scott Jackson. Appendices 5 and 6 were compiled by Amy Ritter and Piyush Singh with the collaboration of Mark Russell and Jeff Wolt. The integrating of the various contributions into a final report and the production of the remaining sections was the work of Russell Jones.

Table of Contents

1.0 Summary	6
2.0 Introduction	7
3.0 Literature Review	9
3.1 Model Validation	9
3.2 Sensitivity Analysis	17
3.3 Comparison of Predictions with Measurements in Model Validation	19
3.4 References	25
4.0 Selection of Experimental Data Sets	31
5.0 Development of Methodology	34
5.1 Procedures for Selecting Input Parameters	34
5.2 Guidance for Performing Calibration Simulations	34
5.3 Statistics	35
5.4 Sensitivity Analysis	35
5.5 Uncertainty Analysis	38
5.6 References	40
6.0 Results of Runoff Simulations	42
7.0 Results of Leaching Simulations	54
8.0 Conclusions and Recommendations	65
Appendices	
1. Summaries of Individual Calibration Papers Cited in Chapter 2	68
2. Guidance Document for Creating PRZM Input Sequences	76
3. Guidance for the Performance of Calibrated Modeling	95
4. Guidance Document for PRZM Sensitivity Analysis Using PB Version 1.2	100
5. Detailed Results of Runoff Simulations	108
6. Detailed Results of Leaching Simulations	362
7. Distributions used in Monte-Carlo Simulations	702
8. Detailed Results of Monte-Carlo Simulations	709

1.0 Summary

The FIFRA Environmental Model Validation Task Force, a collaborative effort of scientists from the crop protection industry and the U.S. Environmental Protection Agency, has compared the results of PRZM3 predictions with measured data collected in 18 different leaching and runoff field studies as part of a process to improve confidence in the results of regulatory modeling.

The work presented in this report shows that PRZM3 provides a reasonable estimate of chemical runoff at the edge of a field. Simulations based on the best choices for input parameters (no conservatism built into input parameters) are generally within an order of magnitude of measured data with better agreement observed both for larger events and for cumulative values over the study period. When the model input parameters are calibrated to improve the hydrology, the fit between predicted and observed improves (results are usually within a factor of three). When conservatism is deliberately introduced into the input pesticide parameters substantial over-prediction of runoff losses occur.

Simulations with PRZM3 show that reasonable estimates of leaching were obtained in homogeneous soils where preferential flow is not significant. PRZM3 usually did a good job of predicting movement of bromide in soil (soil and soil pore-water concentrations were generally within a factor of two of predicted values). For simulations based on the best choices for input parameters (no built in conservatism), predictions of soil concentrations for pesticides were usually within a factor of three and soil pore-water estimates were within a factor of 11. When the model input parameters were calibrated to improve the simulation of hydrology, predicted pesticide concentrations in soil pore-water were usually within a factor of two of measured concentrations. Because of the sensitivity of leaching to degradation rate, the most accurate predictions were obtained with pesticides with relatively slow degradation rates. When conservative assumptions were used to define input pesticide parameters, predictions of pesticide concentrations were usually at least a factor of two greater than predictions made using the best estimate of input parameters without any built in conservatism.

The initial work conducted by different contractors showed the importance of having an SOP which defines the selection of **all** model input parameters. The most satisfactory way to implement regulatory modeling is through the development of a shell which provides all input parameters related to the scenario, with the user providing only the parameters related to the specific pesticide being assessed.

In addition to the results of the comparisons of model predictions with measured values, important products of the task force include a Standard Operating Procedure for conducting site-specific simulations, guidelines for conducting calibration simulations, development of a Plackett-Burman tool for conducting sensitivity analyses with PRZM 3.12, development of a Crystal Ball® Pro Monte-Carlo procedure for assessing the effects of input parameter uncertainty on PRZM 3.12 simulations, and the preparation and publication of data sets which can be used to validate other models in the future.

Recommendations for future work to improve regulatory models include implementation of more sophisticated evapotranspiration routines, allowing for seasonal variation of various model parameters (such as curve numbers, crop cover, and Manning's surface roughness coefficients), better procedures for estimating site-specific degradation rates in surface and subsoils, and improved sorption routines.

2.0 Introduction

Environmental fate models have been used to describe the behavior of crop protection products in the environment since the early 1980's. During the mid-1980's these models started to be used as part of the registration process in various countries to evaluate potential movement of crop protection products to ground and surface water. In recent years, the importance of modeling for evaluating environmental exposure has continued to increase. In 1992 the EPA established a "New Paradigm" for the evaluation of ecological risk, which recommended computer modeling as a cost-effective exposure assessment tool that could help speed up regulatory decision making. Use of modeling was also adopted in the new European Union registration process. As a result of the increased regulatory use of environmental models, several work groups consisting of regulators, industry, and research institutes/environmental consulting firms have been established to develop procedures for the use of these models in estimating environmental concentrations for regulatory risk assessment. One of the first of these groups formed was the FIFRA Exposure Modeling Work Group in the U.S.

Considerable research has been performed over the past 15 years to compare model predictions with experimental results. However, as a result of the limitations of this research, there still exists a general lack of confidence in present-day models and concern remains about their validation status. Reports from the Aquatic Risk Assessment and Mitigation Dialogue Group and the EPA Science Advisory Panel that it spawned have pointed out the critical importance of having confidence in the results of computer models.

Therefore, the FIFRA Environmental Modeling Work Group initiated in 1995 a Model Validation Project, aimed at justifying the use of modeling tools that are needed for risk refinement under EPA's New Paradigm and to address the issues raised by the SAP. This Model Validation Project was funded and conducted by an industry task force (The FIFRA Environmental Model Validation Task Force) composed of scientists from Aventis, BASF, Bayer, Dow AgroSciences, Dupont, FMC, ISK Biosciences, Monsanto, Rohm & Haas, Syngenta, Uniroyal, and Valent. Representatives of the U.S. EPA participated in the technical meetings and served as advisors on all of the subgroups. The work of the group was reviewed by a panel of modeling experts from universities and other government agencies.

Program Objectives

The overall goal of the project was to improve confidence in the primary environmental fate models used for regulatory exposure assessments under FIFRA in the U.S.

For purposes of this project, the term "validation" refers to the characterization of how well the primary models describe the observed fate of pesticides under real world conditions. Rather than requiring model results to be within certain limits of accuracy for a model to be "validated," the intent of this project was to characterize models' strengths and weaknesses and to characterize how close the modeling results were to real-world observations across a wide, but realistic range of input parameters.

As a parallel objective, the project has led to the creation of readily usable data sets and methodology for the calibration and validation of new or revised models in the future. The project has identified which features of the existing primary environmental fate models most require improvement as well as provide a template for the design of future field studies, thus facilitating further model development

Project Outline and Scope

The Model Validation Project has been divided into three phases:

Phase 1: Review of published literature.

Phase 2: Selection of literature and company data sets for modeling simulations.
Development of SOPs for conducting modeling simulations.
Development of statistical procedures for evaluating model performance.
Assembly of literature and company data sets.
Conduct of model simulations.
Comparison of model results to observed data.
Guidance on development of new data sets for model validation.

Phase 3: Programs (if needed) to address model or data deficiencies.

The scope of this validation effort had to be carefully defined to make the task achievable. Exposure assessments were limited initially to movement in the unsaturated zone for leaching assessments and edge-of-field concentrations for runoff assessments. PRZM and GLEAMS were originally chosen as the models to be tested. These two models, currently used within EPA and industry for regulatory exposure assessments, had been previously recommended by FIFRA Environmental Modeling Work Group. Some initial simulations were performed with GLEAMS. However, the decision of USDA not to continue support of GLEAMS as well as difficulties of adequately describing the soil profile in a way that would result in predictions comparable to field data, resulted in the discontinuing the validation work with GLEAMS. Previous validation work with GLEAMS was included in the literature review.

This report describes the program through the first two phases. No decision has been made about the need for or the nature of a phase 3 program. However, some model improvement activities were included as part of the work performed in phase 2

Types of Comparisons

The following three different levels of comparisons of model predictions with field data have been considered in the Model Validation Project:

Cold. This refers to model runs in which no site-specific data (other than weather data) were employed in the model run. Model input parameters were estimated using the typical EPA procedure. This type of assessment provided an estimate of how well current practices work to generate regulatory exposure estimates.

Site-Specific. This refers to model runs in which all available site-specific measured data were employed to define the input parameters. This includes the use of measured on-site soil, hydrologic, and pesticide properties, such as partition coefficients and observed field dissipation half-lives for the site. This provided an estimate of how well the model can be used to describe movement at a specific site.

Calibration. In these model runs, the experimental results were used to refine the values of selected model inputs to provide closer agreement between model predictions and observed field data. These runs determined which parameters require adjustment, evaluated the improvement in fit, and identified model components that may require improvement.

In the first two types of comparisons, the task force developed procedures to prevent the modeler from having access to the field results. In the third type, the modeler had full access to all relevant data.

3.0 Literature Review

3.1 Model Validation

The first phase of the Model Validation Project was to review the existing information on model validation of PRZM and GLEAMS. The primary purpose of this literature review was to assess the quality and quantity of existing information on the validation of PRZM and GLEAMS to determine whether the additional model validation studies are needed. A second purpose of the literature review was to collect information that would be useful in planning future model validation studies. This report summarizes both aspects of this literature review and well as presents the reasons why the FIFRA Exposure Modeling Work Group concluded that more validation research would be useful in improving confidence in models used in regulatory assessments.

Literature on Validation of PRZM and GLEAMS

A literature search identified 35 articles involving the calibration/validation of model simulations with PRZM and/or GLEAMS with measured data. These calibration/validation studies, summarized in Table 3-1, use data from seven countries on three continents as well as a number of different compounds.

Due to the varied nature of the papers and the lack of details for both model predictions and measured results, a detailed systematic comparison of model predictions is not possible. In order to provide qualitative information on model performance, the results of each paper are summarized in Appendix 1. The majority of the papers indicated good agreement between model predictions and measurements or that the models generally predicted more movement than actually occurred. These results over the wide range of conditions reported in the papers lend general support to the use of PRZM and GLEAMS in the regulatory process, especially for predicting leaching.

Some of the deficiencies in the PRZM and GLEAMS models noted in the 35 papers are summarized in Table 3-2. Author's comments on deficiencies were included whether or not such deficiencies were actually reflected in the comparison of predictions with measured data. The larger number of deficiencies listed for the PRZM model is a reflection of the greater use of PRZM in the 35 papers rather than an indication that GLEAMS has less deficiencies. In fact, most of the deficiencies noted in the table are common to both models. Similarly the lack of comments related to runoff is the result of most of the comparisons reported in the papers are for downward movement in the soil profile.

Evaluation of PRZM and GLEAMS Validation Studies

After review of the papers listed in Table 3-1, the FIFRA Exposure Modeling Work Group decided that additional comparisons of field data and model predictions would be useful to supplement existing studies in helping improve confidence in the regulatory use of environmental models for predicting leaching and runoff. The following observations contributed to this decision:

- None of the published studies used the current version of either model (this is especially relevant to PRZM where the runoff routines have been changed significantly).
- Very few of the studies focused on runoff losses (most studies focused on the mobility of crop protection products in the soil profile).
- The number of studies having quantitative validation results is minimal. Since few of the published studies consider model validation the primary purpose of the field experiments, often data sets were not as extensive as would be desirable for model validation.
- Modelers were aware of field results in most of the studies (although in some of the studies where the field results were known, modelers claimed to make no adjustments to the input parameters). Therefore, in these studies the comparisons of model predictions and experimental measurements could be considered calibration since in model validation the modeler should have no knowledge of the field results to prevent biasing the selection of input parameters.

Van den Bosch and Boesten (1995) independently reviewed validation efforts with PRZM, LEACHP, GLEAMS, and PELMO. For both PRZM and GLEAMS, they assessed the quality of the validation efforts in six papers (all of which are included in Table 3-1). They concluded that the validation status of PRZM and GLEAMS was low, especially at concentrations near the European 0.1 µg/L drinking water guideline.

Literature on Validation Research

The literature review also highlighted some areas requiring careful consideration in a model validation study. These include:

Model Validation versus Validation of the Regulatory Modeling Process. In regulatory applications, the purpose is usually to predict the amount or concentration of a compound in runoff water or ground water at a site where extensive research has not been performed. This is in direct contrast to the model developer who usually is trying to fit predictions to existing data obtained from a field experiment. For regulatory applications, the selection of some model parameters (such as soil properties, degradation rates, sorption parameters, or compartment sizes) may not be as straightforward as for the model developer. In regulatory applications, many of the parameters must be obtained from information in data bases or estimated from laboratory studies or studies performed at different locations. Since the selection of model input parameters is usually one of the most important factors affecting the accuracy of predicted results, the validation process must be designed so it is not merely an exercise testing the ability of the modeler to select proper input parameters (Jones and Rao, 1988). However, an incorrect assessment in a regulatory application is equally wrong whether resulting from a poor selection of input parameters or poor model performance. If an incorrect assessment results from the poor choice of input parameters, this is not necessarily an indication of poor model performance. However, if an input parameter to a specific model cannot be selected with sufficient accuracy to ensure satisfactory model predictions, then this model may not be suitable for use in regulatory applications.

Therefore, the validation of a regulatory application of a model must include validation of the procedures for selecting input parameters. This requires that these procedures be exactly described to eliminate (or minimize) the influence of the modeler. As a result validation of a regulatory application requires an additional step beyond the traditional validation process when the model developer validates (or calibrates) the model by comparing its predictions with available experimental data. In the second step both the procedure for selecting input parameters and the resulting model predictions are tested. This two step validation process of testing the model followed by testing of the modeling procedure is necessary to avoid misleading results, since the process of selecting the input variables can compensate for faults in a model.

The importance of the modeling process was shown in a ring test of the PRZM, LEACHP and VARLEACH models (Brown et al., 1996), which demonstrated that differences in judgment even with experienced modelers can affect model predictions (Brown et al., 1996). In this test five modelers were given the same description of a field experiment and were then asked to model the movement of an experimental pesticide to obtain information on the concentration of the pesticide in the soil profile at 220 days after application and in the soil-water at a depth of 1 m. The authors noted that no two sets of predicted results for a given model were exactly the same. This result is not surprising given the wide variation in the assumed values for many of the input parameters. For example all five modelers used five different assumptions about the thickness of the various soil segments. The authors concluded: "The ring test shows that modelling results for the same scenario and model can vary between users. In this case, the variation between five simulations was similar to that associated with the measurements of pesticide behavior in the field. This user-dependence of modeling has not been previously considered, but should be an important component of evaluating model output. For example, much effort is currently targeted at model validation and defining the range of that validity. To date, this effort has involved comparison between a given set of field observations and a single simulation or range of simulations carried out by a single user. Even where predicted results give an acceptably accurate simulation of field behavior, the findings of this ring test suggest that claims of validity will be misleading unless it can be proved that similarly accurate results would be obtained by a number of independent users."

Acceptability of Model Predictions. The acceptability of model predictions compared with field measurements is influenced by use in a regulatory setting. It is critical that regulatory modeling procedures do not significantly under predict the movement of residues into ground or surface water, so that unexpected impacts on the environment do not occur. Model predictions indicating greater movement than what actually occur are not a problem as long as unnecessary restrictions do not result from the risk assessment. The challenge is to develop a modeling process that produces a conservative set of results while minimizing the difference between model predictions and experimental results.

Quantitative Procedures for Comparing Model Predictions with Observed Values. Most validation/calibration comparisons in the studies in Table 3-1 are qualitative, using statements such as the data generally agree with the model predictions. Obtaining statistical descriptions of these comparisons is more difficult; however, a number of papers (for example, Haan et al., 1995; Parrish and Smith, 1990; Walker et al., 1995; Boekhold et al., 1993; Loague et al., 1995) have been published which examine various procedures for quantifying the ability of model predictions to describe observed values. Boekhold et al. (1993) discuss five approaches to assessing model performance: factor f approach, comparison of confidence intervals, comparison of mean values, comparison of variances, and graphical methods. For their validation work with PESTLA they chose the factor f approach, which is based on the capacity index approach described by Parrish and Smith (1990). Walker et al. (1995) present several different indices for expressing the overall fit and descriptions of degradation and movement. Loague et al. (1995) suggest that summary variables that can be obtained from concentration variables include total mass, center of mass, peak concentration, time for a critical concentration to leach to a depth of interest, and depth of the leaching front and advocate the use of root mean square error as a statistical measure of model performance. Comparison of predictions with measurements is discussed in more detail in Section 3-3

The choice of variables for comparison also must consider the regulatory application. For example if a model correctly predicts the amount of a chemical moving to the water table but the timing is off a couple of days, the error in timing makes no difference in a risk assessment. Obviously a model correctly predicting runoff as a function of time within a rainfall event is desirable, but a model that only gets the total loss during an event correct may be adequate in a regulatory application. A model that correctly predicts movement to the water table may be acceptable even if it does not correctly predict soil concentration profiles. However, the fundamental validity of the model processes must be maintained. For example, correctly predicted runoff or leaching losses of crop protection products must be considered irrelevant if water movement is not adequately described. Armstrong et al. (1996) describe a multi-step validation process that considers water movement, tracer movement, and then movement of the specific chemical.

Separation of Modeler from Field Data. To maximize the credibility of a validation exercise that includes the selection of input parameters, the modeler should have no knowledge of the field results. Otherwise, the validation work will probably be characterized as calibration.

Calibration Simulations. If the predictions based on the initial set of input parameters do not provide acceptable agreement, a set of calibration simulations may be performed to help determine whether the source of error is the result of the model or the selection of input parameters. Such calibration should not consist of a simplistic regression of input parameters to minimize difference between observed and predicted values since most water quality models have enough adjustable parameters to fit a limited set of field observations (Haan et al., 1995), but rather a systematic variation of input parameters constrained to feasible ranges. Results of sensitivity analyses (such as those described by Fontaine et al., 1992 and Walker et al., 1995) can be useful tools in performing such simulations. Calibration simulations should normally be performed as a two step process: first calibrating parameters affecting movement of water to optimize the hydrology, then changing chemical-specific properties to best describe chemical movement. Haan et al. (1995) describe a statistical protocol that transforms parameter uncertainty into prediction uncertainty using probability density functions in order to “distinguish a good fit that is based on artificial manipulation of an overparameterized model from a good fit that is based on an accurate description of the processes that control contaminant transport.”

Improving the Quality of Field Studies Used to Validate Models. Additional site-specific information (for example, more detailed soil information, application of a tracer, or soil-specific laboratory sorption and degradation studies) may be useful when a field study is being used to obtain data for model validation. Smith et al. (1990a) review some of the items which should be considered when conducting field studies for model validation. This subject is important when reviewing existing data sets for inclusion in validation studies as well as in the planning of new studies to be used in model validation.

Conclusions

Comparisons of PRZM and GLEAMS predictions with field measurements have been made in a number of studies found in the literature. Most of these studies demonstrated that PRZM can be a useful tool in assessing leaching. Although PRZM predictions in surface soil in the early portions of a study are not particularly useful, its over predicting of residue movement in later stages of an experiment provide conservative assessments suitable for use in estimating potential leaching in regulatory risk assessments. However, because of various limitations of the available literature studies, additional validation research to supplement existing studies would improve the confidence in the runoff and leaching predictions of PRZM and GLEAMS in regulatory applications. This validation research should carefully consider:

- Improving and standardizing the process for selection of input parameters.
- Developing procedures for performing calibration simulations to determine whether differences between model predictions and field measurements are the result of model inaccuracies or the choice of input parameters.
- Devising appropriate procedures for keeping results of field studies from modelers performing simulations to validate model predictions while providing access when calibration simulations are being performed.
- Developing quantitative statistical procedures for comparing model predictions with field measurements.
- Identifying the combinations of soil properties and weather patterns under which the models provide estimates that are sufficiently accurate for use in regulatory decision making.
- Identifying specific areas where each of the models can be improved.
- Identifying the type and quantity of measurements that must be made in field studies to ensure suitability for model calibration and/or validation.

Table 3-1. A partial listing of model validation or calibration studies conducted with GLEAMS and PRZM.

Reference	Models	Locations	Compounds	Soil Types
Barrett, 1995	PRZM	Kansas, U.S.A.	triasulfuron	Las Animas loamy sand
Cai et al., 1993	PRZM	Jiangsu province, China	aldicarb	sandy loam
Carsel et al., 1985	PRZM	New York, U.S.A.	aldicarb	Haven sandy loam
Carsel et al., 1986	PRZM	Florida, U.S.A. Maryland, U.S.A.	metalaxyl	Blanton fine sand Marton fine sandy loam
Dibbern and Pestemer, 1992.	GLEAMS, PRZM, CALF, LEACHM, SESOIL, EQUI	Germany	terbuthylazine	loess soil
Dowd et al, 1993	PRZM	Georgia, U.S.A.	lindane	Cecil (clayey thermic, Typic Kanhapludult
Flori et al., 1993	PRZM	Po Valley, Italy	metamitron, chloridazon, ethofumesate, lenacil	field capacity and wilting point of 33 and 10 vol %
Hegg et al., 1988	PRZM	South Carolina, U.S.A.	aldicarb	Dothan loamy sand
Jones et al., 1983	PRZM	Florida, U.S.A.	aldicarb	sand and fine sands
Jones et al., 1986	PRZM	Arizona, U.S.A. California, U.S.A. Florida, U.S.A. Indiana, U.S.A. Maine, U.S.A. Michigan, U.S.A. Nebraska, U.S.A. New York, U.S.A. North Carolina, U.S.A. South Carolina, U.S.A. Virginia, U.S.A. Washington, U.S.A. Wisconsin, U.S.A.	aldicarb, aldoxycarb aldicarb aldicarb, aldoxycarb aldicarb aldicarb aldicarb aldicarb aldicarb aldicarb, aldoxycarb aldicarb aldicarb, aldoxycarb aldicarb aldicarb, aldoxycarb aldicarb	sandy loam loamy sand and sandy loam sand and fine sands silty clay loam loam sandy loam loamy sand sandy loam sandy loam sandy loam Dothan loamy sand clay loam sandy loam sand, loamy sand, and sandy loam
Jones et al., 1987	PRZM	Nebraska, U.S.A.	aldicarb	loamy sand
Khan and Green, 1988	PRZM	Hawaii, U.S.A.	DBCP	Pauwela clay and Hamakuapoko silty clay
Leonard et al., 1990	GLEAMS	Georgia, U.S.A.	fenamiphos	Cowarts loamy sand

Table 3-1 (continued). A partial listing of model validation or calibration studies conducted with GLEAMS and PRZM.

Reference	Models	Locations	Compounds	Soil Types
Loague, 1992	PRZM	Hawaii, U.S.A.	EDB	Leilehua (Humoxic Tropohumults)
Loague et al., 1989a	PRZM	Hawaii, U.S.A.	EDB	Leilehua (Humoxic Tropohumults) and Wahiawa (Tropheptic Eutrustox)
Loague et al., 1989b	PRZM	Hawaii, U.S.A.	DBCP, EDB, TCP	Leilehua (Humoxic Tropohumults)
Loague, et al., 1995	PRZM	Hawaii, U.S.A.	bromide, chlorpyrifos, fenamiphos	Kawaihapai and Wahiawa volcanic soils
Lorber and Offutt, 1986	PRZM	North Carolina, U.S. A. Wisconsin, U.S.A.	aldicarb	sandy loam loamy sand and sandy loam
Mueller, 1994	PRZM	Sweden (lysimeter)	dichlorprop, bentazon	Lanna clay and Mellby sand
Mueller et al., 1992	GLEAMS, PRZM	Georgia, U.S.A.	alachlor, metribuzin, norflurazon	Dothan loamy sand and Appling sandy loam
Nicholls, 1994	PRZM, CALF	Sweden (lysimeter)	bentazon	Nantuna sand
Parrish et al., 1992	PRZM, AGGR	Georgia, U.S.A.	aldicarb, metolachlor, bromide	loamy sand to sandy loam
Pennell et al 1990	GLEAMS, PRZM, CLMS, MOUSE, LEACHMP	Florida, U.S.A.	aldicarb, bromide	Astatula sand
Perry, 1991	PRZM	Kansas, U.S.A.	atrazine, alachlor, metolachlor, trifluralin, 2,4-D	Eudora silty loam, Eudora sandy loam, and Eudora-Kimo clay
Sadeghi et al., 1995	PRZM	Maryland, U.S.A.	atrazine	Iuka and Hatboro silt loam
Sauer et al., 1990	PRZM	Wisconsin, U.S.A.	atrazine, metolachlor, carbofuran, chlorpyrifos	Plainfield sand
Shirmohammadi and Knisel, 1994	GLEAMS	Sweden (lysimeter)	dichloprop, bentazon	Mellby sand
Shirmohammadi et al., 1987;1989	GLEAMS	Maryland, U.S.A.	atrazine, carbofuran, cyanazine, dicamba, metolachlor, simazine	Matapeake silt loam
Sichani et al., 1991	GLEAMS	Indiana, U.S.A.	alachlor, atrazine, cyanazine, carbofuran, chlorpyrifos	Clermont silt loam

Table 3-1 (continued). A partial listing of model validation or calibration studies conducted with GLEAMS and PRZM.

Reference	Models	Locations	Compounds	Soil Types
Smith, M. C., et al., 1991	GLEAMS, PRZM	Georgia, U.S.A.	atrazine, alachlor, bromide	Lakeland sand
Smith, W. N., et al., 1991	PRZM, LEACHMP	laboratory experiments with intact soil cores	atrazine	sandy loam
Trevisan et al., 1993	PRZM, BAM, LEACHM	Italy	atrazine, metolachlor	loam
Walker et al., 1995	GLEAMS (runoff only); PRZM2, LEACHP, VARLEACH (leaching only)	U.K. (6 locations), Germany (10 locations), France (6 locations), Italy (4 locations)	specific compounds not specified but include alachlor, chloridazon, metribuzin, metsulfuron-methyl, terbutylazine, runoff simulations only with alachlor	various soils, runoff simulations with sandy loam and clay loam
Walker et al., 1996	PRZM, VARLEACH, LEACHP	U.K.	alachlor, atrazine, metribuzin	packed columns of sieved surface soil (75 % sand, 10 % silt, 15 % clay, and 1.91 % organic matter)
Zacharias et al., 1994	GLEAMS, PRZM	Virginia coastal plain, U.S.A.	bromide, atrazine, metolachlor	Suffolk sandy loam

Table 3-2. A summary of selected deficiencies noted by papers summarized in Table 2-1.

Model	Deficiency	References
PRZM	Does not consider preferential flow (including residues in lower layers under predicted due to preferential flow)	Barrett, 1992; Dowd et al., 1993; Loague et al., 1989b; Loague et al., 1995; Nicholls, 1994; Parrish et al., 1992; Perry, 1991; Sadeghi et al., 1995; M. C. Smith et al., 1991; W. N. Smith et al., 1991; Zacharias and Heatwole, 1994
	Overestimates downward movement through soils, especially at later sampling intervals	Carsel et al., 1985; Hegg et al., 1988; Jones et al., 1986; 1987; Loague et al., 1989ab; Loague et al., 1995; Mueller et al., 1992; Parrish et al., 1992; Pennell et al., 1990; Sauer et al., 1990; Trevisan et al., 1993; Walker et al., 1995
	Soil concentration profiles not predicted accurately	Dibbern and Pestemer, 1992; Jones et al., 1986; Loague et al., 1989a; Parrish et al., 1992; Zacharias and Heatwole, 1994
	Difficulties with estimating dispersion accurately (including effect on peak concentrations and the effect of thickness of soil horizon on simulation results)	Barrett, 1992; Flori et al., 1993; Jones et al., 1983; Parrish et al., 1992; Walker et al., 1995
	Under predicts persistence in surface soils	Cai et al., 1993; Jones et al., 1986; Loague et al., 1995; Lorber and Offutt, 1986; Pennell et al., 1990
	Does not consider upward movement due to capillary transport	Loague et al., 1995; Walker et al., 1995
	Estimation routines for evaporation are too simple and inaccurate	Walker et al., 1995; Zacharias and Heatwole, 1994
	Simplicity of degradation description (including independence of degradation rate independent of soil moisture and temperature)	Loague et al., 1995; Mueller et al., 1992; Walker et al., 1995
	soil hydraulics are too simplistic for vadose zone applications or for less porous soils	W. N. Smith et al., 1991
	poor results for uncalibrated simulations of deep leaching (about 20 m)	Loague, 1992
GLEAMS	Does not consider preferential flow (including residues in lower layers under predicted due to preferential flow)	Shirmohammadi and Knisel, 1994; Sichani et al., 1991; M. C. Smith et al., 1991; Zacharias and Heatwole, 1994
	Overestimates downward movement through soils, especially at later sampling intervals	Mueller et al. 1992; Pennell et al., 1990; Shirmohammadi et al., 1987; 1989
	Soil concentration profiles not predicted accurately	Dibbern and Pestemer, 1992; Zacharias and Heatwole, 1994
	Degradation rate is independent of soil moisture	Mueller et al., 1992
	Underestimates surface runoff	Shirmohammadi et al., 1987; 1989
	Model cannot handle partially frozen soil	Shirmohammadi and Knisel, 1994
	Runoff parameters are hard to obtain for soils located outside the U.S.	Walker et al., 1995

3.2 Sensitivity Analysis

One of the goals of FEMVTF is to provide uncertainty bounds for numerical modeling based upon comparison with open literature and industrial data sets (field dissipation, prospective ground water, and runoff studies). Results give both regulators and scientists a measure of how well current environmental fate models can predict real world behavior. An integral part of these tasks is determining the quality of data sets that span a wide range of philosophy for design and conduct of field research. Model sensitivity analysis can be instrumental in deducing the quality of the various data sets by indicating those input parameters needed for accurate model predictions. Such information aids in interpretation of modeling results and may be a necessary precursor to any subsequent statistical comparison of model predicted results to field observations (Haan et al., 1995). Sensitivity analysis can also prove useful for future field study design by providing indications of parameters requiring the greatest accuracy in measurement.

Input Data Sensitivity for Transport Modeling. Multiparametric models, such as transport models, may be sufficiently non-linear in their response to compromise their ability for validation. Sensitivity to input parameter variance can be used to identify sensitive input parameters and, secondarily, can allow for evaluation of model efficacy (ability to produce a desired effect; Addiscott et al., 1995). Sensitivity analysis should be conducted across the full range of likely parameter values, and probably a little further (Addiscott, 1993). In addition to evaluating sensitivity of input parameters, the effect of certain input assumptions which are not parameters per se (for example, selection of layer thickness for a leaching model; Addiscott and Whitmore, 1991) should be considered.

Although transport modelers develop an intuitive sense for sensitive input parameters, this intuitive sense is restricted to the universe of data sets modelers may have evaluated. Statistical approaches to sensitivity analysis can be used to validate modeler intuition and to extend knowledge of model robustness and efficacy across a wider range of input data sets. Unfortunately, limited documentation of input data sensitivity (either intuitive or statistical) exists for transport models.

Empirical Description of Transport Model Sensitivity. Walker et al. (1995) reported the effect single parameter variance on output (total pesticide residues with time, residue distribution with depth, and fraction leached) for a standard set of input data for PRZM2 (as well as LEACHP and VARLEACH). Soil half-life and K_d significantly influenced total soil residues and fraction of residues leached, respectively. A 1.5-fold increase in half-life (133 to 200 days) resulted in a 43 % increase in total soil residues. Variance in K_d from 1.7 to 8 resulted in about 10% greater soil residues. The effect of variance in either half-life or K_f on output was non-linear. When a Freundlich K_f and $1/n$ were used instead of K_d , variance in the Freundlich coefficient ($1/n$) was not judged to be particularly sensitive, but was more important for a lower associated K_f value.

Soil parameters were varied under fixed inputs for weather and chemical properties (Walker et al., 1995). Decreased bulk density (1.5 to 1.0) resulted in 30% greater mean leaching depth. Initial soil moisture content and field capacity were insensitive soil parameters. In contrast, Monte-Carlo analyses with PRZM showed field capacity to be the most sensitive parameter for prediction of leaching for a short-lived (half-life of 30 to 60 days), weakly sorbed ($K_{oc} = 20$ to 40 L kg^{-1}) pesticide (Carsel et al., 1988ab).

Finally, the sensitivity of PRZM2 to the treatment of dispersion was considered an important modeling consideration (Walker et al., 1995). PRZM2 proved particularly sensitive to thickness of layer segments, especially in the surface few centimeters, as this influenced the effect of numerical dispersion. This can be minimized by careful initial evaluation of the effect of segment thickness or, if appropriate site hydrological information are available, through use of an analytical dispersion value. Although results are not reported for PRZM2, with LEACHP set to a 3-cm layer thickness, a change in numerical dispersion from 0 to 20 mm caused an approximate 10 % decrease in peak pesticide concentration without an increase in leaching loss. A further increase to 50 mm resulted in significantly increased leaching loss and a flatter pattern of pesticide residue concentration in the soil profile. Dispersion assumptions have additionally been cited as a critical aspect of PRZM model performance by Carsel et al. (1988ab) and have been addressed in PRZM2 and PRZM3 with the inclusion of the method of characteristics option

that solves the chemical transport equation with elimination or minimization of numerical dispersion (Mullins et al., 1994).

Leonard and Knisel (1988) observed that for GLEAMS leaching, much of the variance in model outcomes for probabilistic modeling with 50 years weather was attributable to rainfall distribution relative to timing of pesticide application. A given large rainfall event on the day of application is sufficient to initiate leaching even for a short-lived molecule. Worst-case rainfall scenarios were suggested as a means to reduce variances. When Koc varied from 100 to 10 L kg⁻¹ for a pesticide with a 60-day half-life, the predicted 50th percentile leaching losses were increased by about 3-fold on a sandy clay loam soil (pesticide leached increased from ca 5 to 15 %). A similar sensitivity analysis on a sand-textured soil showed the same relative magnitude of effect but the absolute amount leached was considerably more significant (pesticide leached increased from ca 12 to 35 %). Leonard et al. (1992) presented similar conclusions regarding annual runoff losses for pesticides as predicted by GLEAMS (that is, rainfall, especially in a short post-application interval has an overriding importance on runoff predictions). These authors simulated moderately to strongly sorbed pesticides (Koc of 100 to 1000) with short soil half-lives (15 days) and also stressed that as half-life increased, sensitivities to inputs such as rainfall might decrease and/or variables related to sediment transport might increase in importance.

Truman and Leonard (1991) investigated GLEAMS predictions of pesticide leaching losses as influenced by environmental fate parameters (surface and subsurface half-lives as well as Koc) for two soil scenarios subjected to the same 50-year pattern of rainfall. As would be expected, increased surface and subsurface half-lives (5 to 30 days and 2.5 to 360 days, respectively) and decreased Koc (10 to 100 L kg⁻¹) increased the amount of predicted leaching loss. Rainfall timing immediately after application was especially important when pesticide half-lives were short. For selected scenarios, potential leaching losses from the root zone increased 2 to 7-fold as subsurface half-life increased by 6-fold.

Sensitivity analysis for pesticide runoff and sediment transport by CREAMS was performed for weakly and strongly sorbed pesticides by Lane and Ferreira (1980). Rainfall was highly significant for a weakly sorbed molecule but not for a strongly sorbed molecule. Application rate and runoff yield as well as application and incorporation efficiency were always significant.

Documentation for GLEAMS similarly indicates sensitive input parameters (Knisel, 1993). Runoff curve numbers (especially CN2) are particularly sensitive parameters that increase in sensitivity as values of CN increase. Porosity and field capacity are additional sensitive parameters governing water flow (leaching versus runoff). When overland flow is segmented to reflect complex slopes, the soil loss ratio (C-factor in USLE) becomes a sensitive parameter. The Koc is the most sensitive pesticide input parameter. For Koc less than 500, surface runoff decreases as Koc decreases because of mobilization below the 0-1 cm surface layer. For Koc greater than 1000, increased Koc shifts pesticide loss from runoff to sediment transport. Soil half-life is sensitive as well. Application rate maybe sensitive depending on the effect of soil half-life and foliar interception to reduce the amount of pesticide available for transport.

Zacharias and Heatwole (1994) used comparisons of bromide and pesticide leaching from uncalibrated and calibrated runs of PRZM and GLEAMS to gain insight as to sensitive components of these transport models. Curve number, field capacity, and wilt point were considered sensitive parameters for both models as were leaf area index for GLEAMS and depth of soil water extraction for PRZM.

Statistical Description of Transport Model Sensitivity. Fontaine et al. (1992) statistically evaluated the effect of input parameter variance for prediction of leaching depth at a fixed total residue concentration when modeled by PRZM within a Monte Carlo shell. Sensitivity analysis was performed for 35 PRZM input parameters when varied over a range appropriate for pre-emergence soybean herbicide use in the midwestern USA. The results were evaluated by both Plackett-Burman (PB) (Plackett and Burman, 1946) and Fourier amplitude sensitivity analysis (FAST). Both statistical tools produced comparable relative sensitivity rankings. The PB analysis proved preferable to FAST because of the much reduced computational intensity of this approach. (PB utilizes a partial factorial design whereas FAST uses Monte Carlo sampling.) Table 4.3-1 shows those inputs with the largest sensitivity

coefficients. The most critical input parameter, key period rainfall, refers to the post application timing of a rainfall event sufficient to initiate leaching and indicates the overriding importance of rainfall/irrigation distribution as a critical PRZM input. PRZM3 will not be markedly different from PRZM in the relative importance of these key input parameters as long as the aforementioned sensitivity to surface layer thickness/dispersion is addressed.

Table 3-3. PRZM input variables of greatest sensitivity for pre-emergence pesticide application to soybean, midwestern USA (Fontaine et al., 1992).

Important for most ranges
Key period rainfall
Pesticide half-life
Detection limit
K _{OC}
Organic carbon fraction
Available water in surface horizon
Important for many ranges
Runoff curve number 3
Runoff curve number 2
Bulk density in surface horizon
Total pesticide applied
Sometimes important
Bulk density in horizon 2
Available water in horizon 2

Cryer and Havens (Cryer and Havens, 1993; Cryer et al., 1994) have utilized PB designs to evaluate sensitive parameters affecting pesticide runoff predictions from GLEAMS and EPICWQ as well as leaching predictions from PRZM2. Analysis of 20 input parameters for EPICWQ used a fixed single year of weather (weather was assumed a priori to be the most significant input class) with a significant effect judged any variance significant at $P < 0.01$ to 0.02 . For chlorpyrifos application to corn in the Midwest, sensitive parameters (ranked) were runoff curve number (CN) after planting, timing of the third application date, timing of the second application date, and the planting date. Similar sensitivity analysis for GLEAMS indicated the sensitive inputs (ranked) were CN, pesticide water solubility, and pesticide soil half-life. (The variance ranges for the analysis were 70 to 95 for the SCS curve number, water solubility of 1 to $6 \mu\text{g L}^{-1}$, and soil half-life of 4 to 70 d; output variables considered were daily maximum and annual fractional runoff and leaching of pesticide, daily maximum and annual water flow rates and erosion mass). The experience of these modelers is that for fixed pesticide properties, ranked sensitivity for runoff inputs is:

weather >>> CN >> all other inputs

and for leaching the dominant sensitive parameters are porosity, field capacity, and hydraulic conductivity (S. Cryer, personal communication).

3.3 Comparison of Predictions with Measurements in Model Validation

In order to validate PRZM3.12 the FEMVTF has studied the varied approaches used to provide a measure of the correlation of model responses to measured data. In general, these measures of correlation or in the converse, residual error, take the form of subjective graphical analyses or objective

statistical evaluations. The following will serve to provide a brief overview of a portion of the available literature concerned with the process of model validation.

Carsel et al. (1986) provide an approach for the evaluation of PRZM1.0 performance based on graphical comparisons of estimated and measured pesticide movement versus depth. PRZM1.0 was calibrated with measured site-specific data and subsequently validated using different hydrologic responses based on long-term climatologic records. Results indicated that PRZM1.0 effectively conserved pesticide mass balance in that integration under the observed and predicted response profiles demonstrated equivalent pesticide mass over time. The primary emphasis of the effort of Jones et al. (1986) was to evaluate the accuracy of the predictions of PRZM1.0. Thirty-four sites were evaluated where observed and predicted maximum leaching depth concentrations of aldicarb or aldoxycarb were compared. The authors evaluated model performance by assessing the leading edge of pesticide movement. The results indicated that the simulated maximum leaching depths generally agreed with the field data. Based on model estimates versus measured data comparisons the authors advocated the use of PRZM1.0 for estimating the magnitude of the depth and timing of the leading edge of pesticide movement but not the ability of the model to predict concentrations. The work of Leonard et al. (1990) illustrates a validation exercise where GLEAMS simulation output was compared to measured data. The model was evaluated without calibration; the effect of the variability of input parameters on model output was not assessed. GLEAMS estimated values for mass of the parent and metabolites of concern were graphically shown to compare favorably with the measured data within the variability of the measured data. Simulated and observed concentrations at depth in the soil at selected dates also closely corresponded. Salo et al. (1994) also provides a validation of the modified CREAMS/GLEAMS model output response compared to measured data from field studies conducted in Finland. Model and measured responses were compared primarily through subjective graphic means. The lack of measured data was discussed particularly in light of the consequence of data deficiencies and the resulting inability to perform statistical analysis. In addition, model sensitivity analysis was performed. Results indicate that model estimates and measured data were in good correlation during the early stages of the study; the correlation declined however over time. The decreasing correlation over time was attributed to the model's inability to account for the interaction of temperature and soil moisture on degradation processes and adsorption/desorption discrepancies.

An extensive discussion regarding the implementation and procedures for model validation efforts have been presented (Smith et al., 1990). The work was specifically focused on PRZM1.0 and RUSTIC (Dean et al, 1989) where model estimates were compared to measure data derived from 4 study sites. The authors note that technical issues that require consideration prior to a validation exercise include a) the need for appropriately data rich benchmark data sets, b) well defined performance and acceptance criteria, c) a recognition of the variability of the observed data and d) well defined parameter estimation procedures. The definition of performance criteria is a recurrent theme found throughout the model validation literature. Defining the performance criteria allows for the relative inaccuracies of model responses versus measured data. This is particularly apparent with regard to concentration and the effect of measuring and estimating parts per billion levels of pesticide in a highly variable environment. Smith et al. (1990) define performance criteria based on the purpose of the modeling analysis. For screening level analyses, a level of accuracy is expected to be within an order of magnitude. For site specific or higher tiered modeling exercises the authors suggest that a factor of 2 - 4 can be sufficient in certain instances but a factor of less than 2 may be appropriate in others.

Smith et al. (1990b) also suggest that objective statistical measures useful for the comparison of model estimates versus measured data include a) paired comparisons of predicted and observed values in space and time, b) integrated comparisons relating to spatially or temporally composited data such as monthly or annual means or totals versus corresponding model estimates and c) comparisons of cumulative frequency distributions of observed data and model predictions in stochastic situations. Statistical measures for paired data and spatially and temporally integrated performance tests include descriptive statistics, error and regression analyses and correlation coefficients. Plots of observed versus predicted values were also advocated as visual indications of agreement. The graphic analysis approaches include a) observed and predicted concentration profiles, b) ranges and medians of integrated observed and predicted data, c) matched predicted and observed integrated values and d)

cumulative distribution functions. The authors note that the variability associated with model input can be addressed by employing Monte Carlo analyses. Uncertainty analysis was discussed within the context of the approaches outlined by Carsel et al. (1988).

Pennell et al. (1990) compared the performance of PRZM 1.0, GLEAMS 1.8.54 and the pesticide component of LEACHM1.0 based on water mass balance, the transport of bromide and the transport and degradation of the reactive solute aldicarb. The indices used to compare and validate the models included the depth to the solute center of mass, the relative mass of aldicarb and metabolites aldicarb sulfoxide and aldicarb sulfone remaining within the root zone and soil profile and solute concentration distributions with soil depth. Objective criteria used to validate and compare the models included root-mean-square error (RMSE), normalized objective function (NOF) and reduced error estimates (REE). Model estimated values were compared to measured field averages for the parameter of interest thereby not specifically considering field level variability. Model evaluation was based on graphical analysis and the objective functions noted previously.

The authors note that it may be unrealistic to expect deterministic pesticide simulation models to accurately predict solute concentration distributions. Comparisons of solute concentration distributions, i.e., accounting for field variability may, as the authors note, be the most rigorous test of a model. Due to intrinsic and extrinsic variability, field measurement of concentration distributions are subject to considerable error. Pennell et al. (1990) conclude ultimately that the ability to validate model predictions of concentration distributions may ultimately be limited by the inability to account for the uncertainty in measured data from within the field.

A description of model evaluation procedures based on graphical displays and statistical criteria has been provided by Loague and Green (1991). The statistical analyses encompass the analysis of residual errors and the differences between observed and predicted values. These include maximum error (ME), root mean square error (RMSE), coefficient of determination (CD), modeling efficiency (EF) and coefficient of residual mass (CRM). In order to demonstrate modeled and measured response correlation, values for ME, RMSE, CD, EF and CRM should approach 0.0, 0.0, 1.0, 1.0 and 0.0, respectively. The authors note however that the statistical analyses defined as above have serious potential limitations due to sample size deficiencies. In addition, the authors advocate the use of graphical analyses including a) comparison of observed and predicted concentration profiles, b) comparisons of ranges and medians of integrated values of predicted and observed data, c) comparison of matched predicted and observed integrated values and d) comparison of cumulative distribution functions for integrated values.

Parrish et al. (1992) employed the factor-of-f technique and measures of goodness of fit to evaluate the predictive capabilities of two models, PRZM and the aggregate model for field transport and transformation (AGGR) versus measured data. The measured responses were compiled over a four-year period conducted in the Dougherty Plain region of Georgia. The leaching characteristics of the pesticides aldicarb, metolachlor and the conserved solute bromide were examined. The factor-of-f technique was based on the outcome of the USEPA Workshop on Field Applicability Testing (USEPA, 1982, unpublished data) where two levels of regulatory use of models were identified. The two defined levels of regulatory use of models include a) screening, where the level of accuracy is anticipated to be within an order of magnitude and b) site specific or higher tiered models where the level of accuracy is required to be within a factor of 2 - 4. The details of the factor-of-f approach are outlined in Parrish and Smith (1990). In addition, based on the approach described by Loague and Green (1991) measures of goodness-of-fit employed included determinations of maximum error (ME), root mean square error (RMSE), coefficient of determination (CD), modeling efficiency (EF) and coefficient of residual mass (CRM).

Allen et al. (1990) in their discussion of the PIRANHA modeling suite suggest that the key to model validation efforts is the appropriate hypothesis test. The authors also advocate paired sample parametric testing procedures. In this case the validity criteria was defined such that the model was deemed "valid" if predictions were within a factor of two of the measured data at least 95% of the time. In addition to the model validity criteria the proposed method of validation included the following: a) determine an

appropriate error measure to test the hypothesis of validity, b) specify the minimum sample size, c) paired samples, d) compute the error measure and test the null hypothesis and e) compare the calculated t-statistic to the appropriate point on the t-distribution.

Zacharias and Heatwole (1994) and Zacharias et al. (1994) employed both subjective (graphical) and objective (parametric statistics) criteria to compare simulated versus measured data. The comparisons of the simulated versus observed data were made considering pesticide concentrations and mass within the root zone, depth of solute peak concentration and center of mass. Additionally, the ratio between observed and predicted data was tested versus the factor of 2 standard. Finally predicted values were categorized with respect to observed median, quartiles and range. Quantitative evaluation of the mass remaining in the root zone and depth of solute center of mass was based on the root mean square error and normalized objective function (NOF). Based on the study GLEAMS and PRZM performed well in predicting pesticide mass, but were less reliable in predicting pesticide concentration distributions in soil.

Typically sample sizes in field studies conducted for testing pesticide transport models are often not large enough to ascertain the nature of the data distribution. The purpose of the Zacharias et al. (1994) study was to present robust quantitative techniques where the validity of the statistical procedures was not dependant upon the assumption of a specific probability model of the population. For model validation this distribution free analysis was based on non-parametric techniques. Specifically, the authors propose a non-parametric approach used to test the factor-of-f approach defined by Parrish and Smith (1990). In addition, non-parametric approaches employing the median value for cases where the mean value of the distribution was not appropriate as the representative value of the distribution of interest were described. These non-parametric methods for analysis of residual error were based on the methods presented by Loague and Green (1991). The non-parametric approaches for the analysis of residual error described by Zacharias et al. (1994b) include the median absolute error (MdAE) calculated rather than the RMSE, the nonparametric coefficient of determination (CD*) and non-parametric modeling efficiency (EF*). Additionally, techniques were described where the assumption of Gaussian distributions were tested.

A stepwise process for the performance of model validation under the auspices of the EU Environmental Research Programme has been described (Melacini and Gunther, 1995). The intent of the study was to provide a uniform basis for the comparison of model performance. The models evaluated included PRZM2.0 the pesticide component of LEACHM and VARLEACH2.0. The authors suggest the use of preliminary subjective graphical analysis where model estimates are compared to measured data. Subsequent objective statistical analyses encompassed tests for evaluation of overall model fit, degradational fit and distributional fit. Measures of the overall fit of model estimates versus measured observations include a) scaled total error, a scaled index based on total observations enabling one to compare results across different studies, b) scaled root mean square error statistics and a model efficiency statistic. The second step of the process entailed the assessment of the degradational fit, i.e., whether the model correctly estimated the correct pesticide mass throughout the soil profile. Statistical measures in this regard include coefficients of residual mass and degradation load, a measure of the weight of the error in prediction relative to the total discrepancy. Finally, measures of the distributional fit of the model, how well or badly the model predicts the distribution of a pesticide in the soil profile was employed. The measures of distributional fit are irrespective of the absolute quantities of the residue involved and include coefficients of determination and shape, a cumulative value test (testing the validity of the simulation of the pesticide movement down the soil profile) and an a determination of the mean depth representing the center of residue mass.

Vanclooster et al. (1996) describe an approach for evaluation multiple models under the auspices of the European COST Action 66 program. The paradigm for the validation procedure using lysimeter data sets entails a calibration or model error minimization step and a subsequent model test versus independent data. Models were evaluated using subjective graphical and objective statistical measures. The graphical analyses include plots of predicted versus measured data, time series plots of predicted and measured data and residual time plots, i.e. plots of model error versus time. Indices of residual error are generally after the methods of Loague and Green (1991) and others and include a) average differences, b) maximum error, c) root mean square error, d) model efficiency, e) coefficient of residual mass and f) coefficient of determination.

Armstrong et al. (1996) advocate a multi-tiered approach to model validation using both graphical and statistical testing. The authors note the pitfalls of any validation effort, namely too few data points collected for parameters of concern, particularly pesticide concentrations in soil and water. Also discussed was the uncertainty associated with model input. The components of the multi-tiered approach include a) parameterization of the model using only independently measured parameters, b) validation of water movement and water content of the soil, c) validation of conserved solute movement, d) validation of pesticides fate in the soil using parameters reflective of independently measured fate information and finally, e) validation of pesticide leaching in terms of comparisons of model predictions with respect to patterns and orders of magnitude of occurrences.

Model input parameter variability in both the spatial and temporal sense has been widely recognized to significant impact model response output. Initial efforts to capture the effects of variability in the temporal sense focused on the consequences of, as an example, the long-term climatologic variation employed by Carsel et al. (1986). Model parameterization reflecting, as examples, adsorption/desorption behavior and degradation were held constant and variability was introduced via the climate data. Thus, the model could be used to infer the behavior of pesticide movements under varied climatologic conditions. The draw back of this approach was that it did not account for the inherent spatial variability of degradation and adsorption/desorption, etc., that is associated with field studies and the behavior of chemicals under environmentally relevant conditions. Subsequent work (Carsel et al., 1988a) employed Monte Carlo techniques to account for the spatial variability noted with respect to soils in field studies. Statistical transformation paradigms particularly the Johnson transformation technique and log transformations were used to convert original random variable, non-normal soil characterization data derived from a national database into normal distributions. Monte Carlo techniques were used to vary field capacity, wilting point and organic matter, soil hydrologic group, weather year and degradation rate (triangular distribution). While no validation exercise was performed, results pertinent to the current discussion demonstrated cumulative probability distribution functions for annual pesticide movement and the influence of model input variability on model response estimates.

A stochastic approach for the PELMO model has been developed and termed MCPELMO (Klein et al., 1999). MCPELMO is based on the deterministic PELMO model but in addition has an incorporated shell allowing for the stochastic simulations for a number of geographically diverse regions. The stochastic efforts entailed climatic variations where nine locations and 30 years of measured weather data were employed. The authors concluded that the temporal variation afforded by the weather data had a greater impact on low mobility pesticides than high mobility pesticides. The authors also suggests that spatial variability in soil may be more influential than the temporal variability on model estimates.

Boekhold et al. (1993) present a validation methodology for the PESTicide Leaching and Accumulation in soil model (PESTLA) that includes both statistical and graphic analyses. The factor-f-approach (Parrish and Smith, 1990) was again used. The factor-of-f approach accounted for the aspect of uncertainty of size f around the predicted value. Thus the model is considered reflective of the relevant environmental situation when measurements falls within the range of acceptable model predictions as defined by the size of the f factor. The authors also discuss an approach based not on the derivation of the predicted mean from a deterministic simulation but on the estimate of the predicted mean and predicted standard deviation of model output based on a stochastic sampling approach for model input variables. The utility of the Monte Carlo analysis was that it provided a quantitative measure of model output uncertainty as a function of model input variability or uncertainty. In addition to typical subjective and objective measures, the authors advocate the use of comparisons of cumulative probability distribution functions (cpdf) of observed and predicted data.

As a follow-up to the previous study, van den Bosch and Boesten (1994) provide an approach for validation of the PESTLA model where both graphical and objective statistics were utilized. Because measured data reflecting the endpoint of concern i.e. groundwater concentrations of ethoprophos or bentazone were scant, the authors turned to an alternate metric for use in the validation process. The metric selected was center of pesticide mass as a function of soil core depth. The underlying assumption was that this measured value would reflect ultimate groundwater concentrations providing that the travel time between the center of mass and the water table was minimal.

The peak concentration values were log transformed and the average and standard deviation of the average were calculated. The confidence intervals around the average of the measured peak concentration of one or two times the standard deviation was used to provide a measure of the uncertainty associated with measured data. The resulting interval boundaries were transformed back to a normal distribution, to give the range of uncertainty in the measurements, used for the statistical test. The factor-of-f approach was applied where $f = 2$ and $f = 5$.

The model was tested across 11 cases; 6 cases for ethoprophos and 5 cases were based on the bentazone data set. Validation results were dependant upon time and the factor-of-f used to validate the model. In seven cases the model was validated regardless of the uncertainty ($f = 2$ or 5) allowed around the predicted value. In one case during the later stages of the study the model could not be validated regardless of the uncertainty factor selected. In the remaining cases the model was validated only when high uncertainty was allowed around the predicted value.

Eckhardt and Wagenet (1996) evaluated the consequences of the inherent variability in soil hydrology and chemical applications and the uncertainty of measurements of soil and chemical properties on the leaching potential of atrazine. The model employed was the pesticide component of LEACHM (Hutson and Wagenet, 1992). The methodology included a model calibration step where model and measured response differences over a 120-day cropping period were minimized via calibration of water flow, as influenced by evapotranspiration and plant root water uptake, and adjustments of the first-order degradation rate of atrazine.

Following the calibration step, the impact on model output response due to the uncertainty associated with two critical transport parameters a) unsaturated hydraulic conductivity and b) atrazine degradation rates were evaluated. Additionally, the effect on model output response due to the uncertainty associated with the spatially heterogeneity of pesticide applications was also evaluated. The uncertainty analysis was conducted underlain by 22 years of weather data measured at the test site. The effects of spatial variability in the hydraulic conductivity of the soil and the uncertainty of degradation rates below the root zone were represented through discrete sampling from probability density functions. The probability density functions from which the discrete samples for hydraulic conductivity and pesticide degradation rate were selected were defined based on empirical data. The results clearly indicate that the uncertainty associated with hydraulic conductivity, pesticide degradation rate and application rates influenced model results.

Haan et al. (1995) provide a discussion regarding the evaluation of model performance in a situation where there are no observed data on the quantities being modeled to assist in input parameter estimation or model calibration. The procedures employed by Haan et al. (1995) include a) the conduct of a sensitivity analysis on model input parameters, b) generation of probability distributions functions (pdf) of input parameters, c) generate probability distributions of model output and d) use the output probability distributions to assess the model. The authors advocate that the use of the measured data be held in reserve such that the measured data do not enter the parameter estimation process. The authors note that the measured responses are also subject to uncertainty. The uncertainty can also be quantified using probability density functions. The overlap of the model output and measured response pdfs can if available be utilized to evaluate model performance.

3.4 References

- Addiscott, T. M. 1993. Simulation Modeling and Soil Behavior. *Geoderma* 60:15-40.
- Addiscott, T. M., J. Smith, and N. Bradbury. 1995. Critical Evaluation of Models and their Parameters. *J. Environ. Qual.* 24:803-807.
- Addiscott, T. M., and A. P. Whitmore. 1991. Simulation of Solute Leaching in Soils of Different Permeabilities. *Soil Use Manage.* 7:94-102.
- Allen, W. B., Jr., M. C. Barber, S. L. Bird, L. A. Burns, J. M. Cheplick, D. R. Hartel, C. A. Kittner, F. L. Mayer, Jr., L. A. Suarez, and S. E. Wooten. 1990. Pesticide and Industrial Chemical Risk Analysis and Hazard Assessment Version 2.0. Environmental Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Athens GA 30613-0801.
- Armstrong, A. C., A. M. Portwood, P. B. Leeds-Harrison, G. L. Harris, and J. A. Catt. 1996. The Validation of Pesticide Leaching Models. *Pestic. Sci.* 48:47-55.
- Barrett, M. R. 1995. Using PRZM Modeling to Compare Leaching Potential of Wheat Herbicides. In: M. L. Leng, E. M. K. Leovey, and P. L. Zubkoff, eds. *Agrochemical Environmental Fate State of the Art*. CRC Lewis Publishers, New York.
- Boekhold, A. E., H. van den Bosch, J. J. T. I. Boesten, M. Leistra, F. A. Swartjes, and A. M. A. van der Linden. 1993. Validation of the PESTLA model: Definitions, objectives and procedure. RIVM report no. 715802002. National Institute of Public Health and Environmental Protection, Bilthoven, the Netherlands.
- van den Bosch, H. and J. J. T. I. Boesten. 1994. Validation of the PESTLA Model: Field Test for Leaching of Two Pesticides in a Humic Sandy Soil in Vredepeel (The Netherlands). DLO Winard Staring Centre Report 82, Wageningen, The Netherlands.
- van den Bosch, H., and J. J. T. I. Boesten. 1995. Validation of the PESTLA Model: Evaluation of the Validation Statuses of the Pesticide Leaching Models PRZM-1, LEACHP, GLEAMS, and PELMO. DLO Winard Staring Centre Report 83, Wageningen, The Netherlands, 60 pp.
- Brown, C. D., U. Baer, P. Günther, M. Trevisan, and A. Walker. 1996. Ring Test with the Models LEACHP, PRZM-2 and VARLEACH: Variability between Model Users in Prediction of Pesticide Leaching Using a Standard Data Set. *Pestic. Sci.* 47:249-258.
- Burns, L. A. 1989. Exposure Analysis Modeling System: User's Guide for EXAMS II Version 2.94. EPA/600/3-89/084, U.S. Environmental Protection Agency, Environmental Research Laboratory, Athens, GA, 99 pp.
- Cai D., Xiang F., Jiang X., Zhu Z., Hua X., and Dai Z. 1993. Fate of Aldicarb in the Vadose Zone Beneath a Cotton Field. *J. Contam. Hydrol.* 14:129-142.
- Carsel, R. F., L. A. Mulkey, M. N. Lorber, and L. B. Baskin. 1985. The Pesticide Root Zone Model (PRZM): A Procedure for Evaluating Pesticide Leaching Threats to Groundwater. *Ecol. Model.* 30:49-69.
- Carsel, R. F., W. B. Nixon, and L. G. Ballantine. 1986. Comparisons of Pesticide Root Zone Model Predictions with Observed Concentrations for the Tobacco Pesticide Metalaxyl in Unsaturated Zone Soils. *Environ. Toxicol. Chem.* 5:345-353.
- Carsel, R. F., R. S. Parrish, R. L. Jones, J. L. Hansen, and R. L. Lamb. 1988a. Characterizing the Uncertainty of Pesticide Leaching in Agricultural Soils. *J. Contam. Hydrol.* 2:111-124.

Carsel, R. F., R. S. Parrish, R. L. Jones, J. L. Hansen, R. L. Lamb, and M. P. Anderson. 1988b. A Simulation Procedure for Groundwater Quality Assessments of Pesticides. *J. Contam. Hydrol.* 2:125-138.

Cryer, S. A., and P. L. Havens. 1993. The Role of Sensitivity Analysis in Modeling Potential Environmental Concentrations Resulting from Pesticide Laden Agricultural Runoff. DowElanco/Purdue Colloquium, 25—26 May 1993, Purdue University, West Lafayette, IN.

Cryer, S. A., P. L. Havens, R. A. Swearington, and P. G. Nordstrom. 1994. Addressing Regional Runoff Variability using a Binned Interval Approach. 748 *In Abstr. 8th Intrnl. Congr. Pest. Chem.*, 4—9 Jul, 1994, Washington, D. C.

Dean, J. D., P. S. Huyakorn, A. S. Donigian, Jr., K. A. Voos, R. W. Schanz, Y. J. Meeks, and R.F. Carsel. 1989. Risk of Unsaturated/Saturated Transport and Transformation of Chemical Concentrations (RUSTIC). I. EPA/600/3-89/048a and II. EPA/600/3-89/048b. USEPA, Athens, GA.

Dibbern, V. H., and W. Pestemer. 1992. Anwendbarkeit von Simulationsmodellen zum Einwaschungsverhalten von Pflanzenschutzmitteln im Boden. *Nachrichtenbl. Deut. Pflanzenschutzd.* 44(6):134-143.

Dowd, J. F. P. B. Bush, D. G. Neary, J. W. Taylor, and Y. C. Berisford. 1993. Modeling Pesticide Movement in Forested Watersheds: Use of PRZM for Evaluating Pesticide Options in Loblolly Pine Stand Management. *Environ. Toxicol. Chem.* 12:429-439.

Eckhardt, D. A. V. and R. J. Wagenet. 1996. Estimation of the Potential for Atrazine Transport in a Silt Loam Soil. *ACS Symposium Series No. 630.*

Enfield, C. G., R. F. Carsel, S. Z. Cohen, T. Phan, and D. M. Walters. 1982. Approximating Pollutant Transport to Ground Water. *Ground Water* 20:711-722.

Fisher, L. J. 1992. Decisions on the Ecological, Fate, and Effects Task Force. Memorandum to D Campt. Office of Pesticide Programs, U.S. Environmental Protection Agency, Washington, D.C.

Fontaine, D. D., P. L. Havens, G. E. Blau, and P. M. Tillotson. 1992. The Role of Sensitivity Analysis in Groundwater Risk Modeling for Pesticides. *Weed Technol.* 6:716-724.

Flori, P., A. Rambaldi, P. Sgarbi, G. Malucelli, and P. Musacci. 1993. Herbicide Mobility and Persistence in Soil: Comparison Between Experimental Data and Predictions by Means of the Pesticide Root Zone Model (PRZM) for the Validation of Mathematical Model. In: A. A. M. Del Re et al., eds. *IX Symposium Pesticide Chemistry, Mobility and Degradation of Xenobiotics*, pp. 105-114.

Haan, C. T., B. Allred, D. E. Storm, G. J. Sabbagh, and S. Prabhu. 1995. Statistical Procedure for Evaluating Hydrologic/Water Quality Models. *Trans. ASAE* 38(3):725-733.

Hedden. 1986. Example Field Testing of Soil Fate and Transport Model PRZM, Dougherty Plain, Georgia. In: S. C. Hern and S. M. Melancon, eds., *Vadose Zone Modeling of Organic Pollutants*. Lewis Publishers: Chelsea, MI, pp. 81-101.

Hegg, R. O. W. H. Shelley, R. L. Jones, and R. R. Romine. 1998. Movement and Degradation of Aldicarb Residues in South Carolina Loamy Sand Soil. *Agric. Ecosystems Environ.* 20:303-315.

Hollis, J. M., S. H. Hallett, and C. A. Keay. 1993. The Development and Application of an Integrated Database for Modelling the Environmental Fate of Herbicides. In: *Proceedings 1993 Brighton Crop Protection Conference-Weeds*, Vol. 3. British Crop Protection Council, Farnham, U.K. pp. 1355-1364.

Hutson, J. L. and R. J. Wagenet. 1992. LEACHM: Leaching Estimation and Chemistry Model. Department of Soil, Crop and Atmospheric Sciences, Research Series No.: 92-3.

Jaeger, R. B. 1995. Transmittal of the Final Report of the Joint Science Advisory Board (SAB) and the FIFRA Scientific Advisory Panel on the Aquatic Dialogue Group Report: Pesticide Risk Assessment & Mitigation; Meeting held July 17, 1995. Memorandum to D M Barolo. Office of Pesticide Programs, U.S. Environmental Protection Agency, Washington, D.C.

Jones, R. L., and P. S. C. Rao. 1988. Reflections on Validation and Applications of Unsaturated Zone Models. In: P. J. Wierenga and D. Bachelet, eds. Validation of Flow and Transport Models for the Unsaturated Zone: Conference Proceedings; May 23-26, 1988, Ruidoso, New Mexico. Research Report 88-SS-04. Department of Agronomy and Horticulture, New Mexico State University, Las Cruces, NM. pp. 197-205.

Jones, R. L., P. S. C. Rao, and A. G. Hornsby. 1983. Fate of Aldicarb in Florida Citrus Soil: 2. Model Evaluation. In: D. M. Nielson and M. Curl, eds. Proceedings of Characterization and Monitoring of the Vadose (Unsaturated) Zone, Las Vegas, December 8-10, 1983, National Well Water Association, Worthington, OH, pp. 959-978.

Jones, R. L., G. W. Black, and T. L. Estes. 1986. Comparison of Computer Model Predictions with Unsaturated Zone Field Data for Aldicarb and Aldoxycarb. Environ. Toxicol. Chem. 5:1027-1037.
Jones, R. L., S. D. Kirkland, and E. L. Chancey. 1987. Measurement of the Environmental Fate of Aldicarb Residues in a Nebraska Sand Hills Soil. Appl. Agricul. Res. 2:177-182.

Khan, M. A., and R. E. Green. 1988. Use of the Pesticide Root Zone Model to Assess DBCP Leaching in Hawaii. Research Series 054, College of Tropical Agriculture and Human Resources, University of Hawaii, 42 pp.

Klein, M. 1995. PELMO Pesticide Leaching Model. Version 2.01. Fraunhofer-Institut für Umweltchemie und Ökotoxikologie, Schmallenburg, Germany.

Klein, M. 1999. Monte Carlo Analysis Using Pesticide Fate Models. J. Pesticide. Sci. 24:55-59.

Knisel, W. G. 1993. GLEAMS: Groundwater Loading Effects of Agricultural Management Systems. Version 2.10. Univ. Georgia Coastal Plain Exp. Sta., Tifton.

Lane, L. J., and V. A. Ferreira. 1980. Sensitivity Analysis. Chapter 6 In W. G. Knisel (ed.), CREAMS: A Field-Scale Model for Chemicals, Runoff, and Erosion from Agricultural Management Systems. USDA Cons. Res. Rep. No. 26.

Leonard, R. A., and W. G. Knisel. 1988. Evaluating Groundwater Contamination Potential from Herbicide Use. Weed Technol. 2:207-216.

Leonard, R. A., W. G. Knisel, and D. A. Still. 1987. GLEAMS; Groundwater Loading Effects of Agricultural Management Systems. Trans. ASAE, paper no. 86-2511.

Leonard, R. A., W. G. Knisel, F. M. Davis, and A. W. Johnson. 1990. Validating GLEAMS with Field Data for Fenamiphos and its Metabolites. J. Irrigation and Drainage Eng. 116:24-35.

Leonard, R. A., C. C. Truman, W. G. Knisel, and F. M. Davis. 1992. Pesticide Runoff Simulations: Annual Means vs. Event Extremes. Weed Technol. 6:725-730.

Loague, K. 1992. Simulation of Organic Chemical Movement in Hawaii Soils with PRZM: 3. Calibration. Pacific Sci. 46:353-373.

Loague, K., and R. E. Green. 1991. Statistical and Graphical Methods for Evaluating Solute Transport Models: Overview and Application. J. of Contaminant Hydrology 7:51-73.

- Loague, K. M., R. E. Green, C. C. K. Liu, and T. C. Liang. 1989a. Simulation of Organic Chemical Movement in Hawaii Soils with PRZM: Preliminary Results for Ethylene Dibromide. *Pacific Sci.* 43:67-95.
- Loague, K., T. W. Giambelluca, R. E. Green, C. C. K. Liu, T. C. Liang, and D. S. Oki. 1989b. Simulation of Organic Chemical Movement in Hawaii Soils with PRZM: 2. Predicting Deep Penetration of DBCP, EDB, and TCP. *Pacific Sci.* 43:362-383.
- Loague, K., R. N. Miyahira, R. E. Green, D. S. Oki, T. W. Giambelluca, and R. C. Schneider. 1995. Chemical Leaching Near the Waiawa Shaft, Oahu, Hawaii: 2. Modeling Results. *Ground Water* 33:124-138.
- Lorber, M. N., and C. K. Offutt. 1986. A Method for the Assessment of Ground Water Contamination Potential Using a Pesticide Root Zone Model (PRZM) for the Unsaturated Zone. In: W. Y. Garner, R. C. Honeycutt, and H. N. Nigg, eds. *Evaluation of Pesticides in Groundwater*. ACS Symp. Ser. 315:342-365.
- Melacini, A. and P. Gunther. 1995. Preliminary Methods for Assessing and Ranking Model Performance in the Prediction of Pesticide Soil Residues. In: *Evaluation and Improvement of Mathematical Models of Pesticide Mobility in Soils and Assessment of Their Potential to predict Contamination of Water Systems*. Eds. Walker, A., Calvet, R., Del Re, A.A.M, Pestemer, W. and J.M. Hollis. Blackwell Wissenschafts-Verlag GmbH, Berlin.
- Mueller, T. C. 1994. Comparison of PRZM Computer Model Predictions with Field Lysimeter Data for Dichlorprop and Bentazon Leaching. *J. Environ. Sci. Health A29(6)*:1183-1195.
- Mueller, T. C., R. E. Jones, P. B. Bush, and P. A. Banks. 1992. Comparison of PRZM and GLEAMS Computer Model Predictions with Field Data for Alachlor, Metribuzin and Norflurazon Leaching. *Environ. Toxicol. Chem.* 11:427-436.
- Mullins, J. A., R. F. Carsel, J. E. Carsel, J. E. Scarbrough, and A. M. Ivery. 1994. PRZM-2, a Model Predicting Pesticide Fate in the Crop Root and Unsaturated Soil Zones: Users Manual for Release 2.0. USEPA Environmental Research laboratory, Athens, GA.
- Nicholls, P. H. 1994. Simulation of the Movement of Bentazon in Soils Using the CALF and PRZM Models. *J. Environ. Sci. Health A29(6)*:1157-1166.
- Parrish, R. S., and C. N. Smith. 1990. A Method for Testing whether Model Predictions Fall within a Prescribed Factor of True Values, with an Application to Pesticide Leaching. *Ecol. Modelling* 51:59-72.
- Parrish, R. S., C. N. Smith, and F. K. Fong. 1992. Tests of the Pesticide Root Zone Model and the Aggregate Model for Transport and Transformation of Aldicarb, Metolachlor, and Bromide. *J. Environ. Qual.* 21:685-697.
- Pennell, K. D., A. G. Hornsby, R. E. Jessup, and P. S. C. Rao. 1990. Evaluation of Five Simulation Models for Predicting Aldicarb and Bromide Behavior Under Field Conditions. *Water Resources Research* 26:2679-2693.
- Perry, C. A. 1991. Observed and Simulated Distribution of Selected Herbicides in Silty Loam, Sandy Loam, and Clay Soil Profiles near Topeka, Kansas. U. S. Geological Survey Water-Resources Investigations Report 91-4017, 71 pp.
- Plackett, R. L. and J. P. Burman. 1946. The Design of Optimum Multifactorial Experiments. *Biometrika* 33:305-325.
- Sadeghi, A. M., A. R. Isensee, and A. Shirmohammadi. 1995. Atrazine Movement in Soil: Comparison of Field Observations and PRZM Simulations. *J. Soil Contam.* 4(2):151-161.

- Salo, S., M. Posch, and S. Rekolainen. 1994. Testing the Modified CREAMS/GLEAMS Model for Pesticide Concentration in Soil. *Agricultural Science in Finland* 3:59-67.
- Sauer, T. J., K. J. Fermanich, and T. C. Daniel. 1990. Comparison of the Pesticide Root Zone Model Simulated and Measured Pesticide Mobility under Two Tillage Systems. *J. Environ. Qual.* 19:727-734.
- SETAC. 1994. Final Report of the Aquatic Risk Assessment and Mitigation Dialogue Group. Pensacola, FL: Society of Environmental Toxicology and Chemistry and SETAC Foundation for Environmental Education, 220 pp.
- Shirmohammadi, A., and W. G. Knisel. 1994. Evaluation of the GLEAMS Model for Pesticide Leaching in Sweden. *J. Environ. Sci. Heath*, A29(6):1167-1182.
- Shirmohammadi, A., W. L. Magette, R. B. Brinsfield, and K. Staver. 1987. Ground Water Loading of Agrochemicals in the Coastal Plain Region. Paper No. 87-2632 presented at 1987 Winter Meeting of the American Society of Agricultural Engineers, December 15-18, Chicago.
- Shirmohammadi, A., W. L. Magette, R. B. Brinsfield, and K. Staver. 1989. Ground Water Loadings of Pesticides in the Atlantic Coastal Plain. *Ground Water Monitoring Review* 9(4):141-148.
- Sichani, S. A., B. A. Engel, E. J. Monke, J. D. Eigel, and E. J. Kladvko. 1991. Validating GLEAMS with Pesticide Field Data on a Clermont Silt Loam Soil. *Trans. ASAE* 34(4):1732-1737.
- Smith C. N., R. S. Parrish, and D. S. Brown. 1990a. Conducting Field Studies for Testing Pesticide Leaching Models. *Intern. J. Environ. Anal. Chem.* 39:3-21.
- Smith, C. N., R. S. Parrish, R. F. Carsel, A. S. Donigian, and J. M. Cheplick. (1990b). Validation Status of Pesticide Leaching and Groundwater Transport Model (PRZM). Environmental Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency. Athens, Georgia 30613-7799.
- Smith, M. C., A. B. Bottcher, K. L. Campbell, and D. L. Thomas. 1991. Field Testing and Comparison of the PRZM and GLEAMS Models. *Trans. ASAE* 34(3):838-847.
- Smith, W. N., S. O. Prasher, and S. F. Barrington. 1991. Evaluation of PRZM and LEACHMP on Intact Soil Columns. *Trans. ASAE* 34(6):2413-2420.
- Trevisan, M., E. Capri, and A. A. M. Del Re. 1993. Pesticide Soil Transport Models: Model Comparisons and Field Evaluation. *Toxicol. Environ. Chem.* 40:71-81.
- Truman, C. C., and R. A. Leonard. 1991. Effects of Pesticide, Soil, and Rainfall Characteristics on Potential Pesticide Loss by Percolation — A GLEAMS Simulation. *Trans. ASAE* 34:2461-2468.
- Vanclooster, M., J. J. T. I. Boesten, and M. Trevisan. 1996. Procedure for Evaluating and Comparing Pesticide Models in the Framework of Cost Action 66.
- Walker, A., R. Calvet, A. A. M. Del Re, W. Pestemer, and J. M. Hollis. 1995. Evaluation and improvement of mathematical models of pesticide mobility in soils and assessment of their potential to predict contamination of water systems. Blackwell Wissenschafts-Verlag GmbH, Berlin, 115 pp.
- Walker, A., S. J. Welch, A. Melacini, Y. H. Moon. 1996. Evaluation of Three Pesticide Leaching Models with Experimental Data for Alachlor, Atrazine and Metribuzin. *Weed Research* 36:37-47.
- Zacharias, S., and C. D. Heatwole. 1994. Evaluation of GLEAMS and PRZM for Predicting Pesticide Leaching under Field Conditions. *Trans. ASAE* 37(2):439-451.

Zacharias, S. Heatwole, C. D. and C. W. Coakley. 1994. Comparison of Quantitative Techniques for Pesticide Model Validation. American Society of Agricultural Engineers, Paper No.: 93-2506.

4.0 Selection of Experimental Data Sets

Nine runoff and nine leaching data sets were selected for comparisons with model predictions. These data sets are summarized in Tables 4-1 and 4-2. More complete information on each data set is provided in Appendices 5 and 6.

The first step in the selection of data sets was to prepare criteria for determining “ideal” and “acceptable” data sets, based on the data requirements as identified in the guidance documents for selection input parameters for GLEAMS and PRZM as well as the consensus opinion of the task force scientists and advisors. Data sets fitting the “ideal” criteria were used when available to the task force but other data sets were also used. For both leaching and runoff studies, the criteria resulted in heavy emphasis on industry data sets conducted within the U.S. to fulfill registration needs. In addition data sets were obtained from the literature and from non-industry researchers. Studies excluded from consideration were those conducted outside the U.S. (due to the limitations of the SOP for selection of input parameters), lysimeter studies, and field dissipation studies (with sampling less than 1 m).

Table 4-1. The nine data sets used in the runoff simulations.

Data Set	Area (ha)	Slope (%)	Soil Type	Crop	Application Method ¹	Half-Life ² (d)	Kd ² (ml/g)
GA1R	3.64	3.5	Loamy sand	Cotton	Aerial (L)	6	4
GA2R	3.04	3-5.5	Sandy loam	Sweet Corn	Foliar (L)	8	0.43
IA2R	7.0	4.3	Silt loam	Corn	T-Band, foliar, and broadcast (G, L)	30	121
IA3R	0.065	5.6	Silt loam	Corn	T-Band, foliar, and broadcast (G, L)	30	121
IA4R	1.21	2.9	Silt clay loam	Corn	T-Band (G)	52	4200
					T-Band (G)	121	12
IA5R	0.065	2.8	Silt clay loam	Corn	T-Band (G)	52	3200
					T-Band (G)	121	10
KY2R	0.065	4.2-5.2	Silt loam	Corn	T-Band (G) ³	52	2200
					T-Band (G) ⁴	121	6
MD1R	0.50	2.0	Silt loam	Corn	Ground spray (L)	165	0.6
MS1R	2.1	0.25	VF sandy loam	Cotton	Foliar (L)	5.7	4.1

1. Formulation type given in parentheses (G = Granular, L = Liquid)

2. Level 2 half-life and Kd values for surface horizons reflect the values provided by the registrants.

3. Applied to plot 1 and 2

4. Applied to plot 3

Table 4-2. The nine data sets used in the leaching simulations.

Data Set	Soil Type	Hydrologic Group	Crop	Application Rate (kg ai/ha)	Half-Life¹ (d)	Kd¹ (ml/g)
CA1L	Loamy sand	B	Alfalfa	0.848	115	0.186
FL1L	Sand	A	Citrus	5.6	17.5	0.088
GA1L	Fine sandy loam	A	Sweet corn	12.6	8	0.176
GA2L	Mixture, ~loamy sand	B / C	Peanuts	2.49 2.35	69 20	0.42 0.11
KS1L	Sandy loam	C	Wheat	0.074	21	0.50
NC1L	Loamy sand	A	Soybean	0.56	100	0.361
NC2L	Loamy sand	A	Soybean	0.14	47	0.425
NC3L	Loamy sand	A	Cotton	0.14	14	0.091
NC4L	Loamy sand	A	Soybean	0.56	365	0.35

1. Site-specific half-life and Kd values for surface horizons reflect the values provided by the registrants.

Ideal Runoff Data Sets

- Complete site-specific weather data covering the period of the study and including daily precipitation, daily temperature, and pan evaporation.
- Site dimensions, slope, and characterization of the occurrence of non-sheet flow.
- Calculation of curve numbers possible for each quantified runoff event.
- Site-specific soil physicochemical properties and profile description.
- Information on time, rate, and method of pesticide application.
- Site-specific laboratory measurements of soil half-life and Kd.
- Foliar decay rates for foliarly-applied materials.
- Daily runoff volume and sediment yield data.
- Water- and sediment-phase pesticide concentrations measured with verifiable methodology and sensitivity.
- Studies conducted and documented by a verifiable standard for QA/QC.

Acceptable Runoff Data Sets

- Spatially and temporally contemporaneous weather data available from a NOAA site.
- Natural field drainage channels known and described from a soil survey map.
- Representative curve numbers obtained from a data base using soil hydrologic group, soil texture, management practice, and crop.
- Measured soil texture and organic carbon for the surface horizon.
- Number and thickness of soil horizons obtained from a data base.
- A measured soil half-life and sorption coefficient.
- Documented management practices and timings of critical events.
- Daily runoff volume and sediment yield data.
- Total pesticide concentration via an acceptable method.
- Peer-reviewed data and interpretations.

Ideal Leaching Data Sets

- Complete site-specific weather data covering the period of the study and including daily precipitation, daily temperature, and pan evaporation.
- Edge-of-field runoff yield or temporal soil moisture with depth.
- Data for a conservative tracer of water flow.
- Site-specific soil physicochemical properties and profile description.

- Information on time, rate, and method of pesticide application.
- Site-specific laboratory measurements of soil half-life and K_d or predictions from laboratory data on the basis of site-specific soil properties.
- Output data including tracer concentrations and/or soil moisture measurements with depth and time.
- Vadose zone measurements in replicate and at several depths providing pesticide and tracer concentrations in soil water and total soil residue concentrations.
- Verifiable analytical methodology with adequate method sensitivity.
- Documentation of sampling design and suction lysimeter placement.
- Data sets with demonstrated pesticide detects in ground water.
- Studies conducted and documented by a verifiable standard for QA/QC.

Acceptable Leaching Data Sets

- Spatially and temporally contemporaneous weather data available from a NOAA site.
- Representative curve numbers obtained from a data base using soil hydrologic group, soil texture, management practice, and crop.
- Measured soil texture and organic carbon by horizon.
- Number and thickness of soil horizons obtained from a data base.
- A measured soil half-life and sorption coefficient.
- Documented management practices and timings of critical events.
- Estimated soil water content on the basis of weather and soil physicochemical properties.
- Output data including total pesticide and tracer concentrations with depth and time determined via an acceptable method.
- Peer-reviewed data and interpretations.

5.0 Development of Methodology.

The literature review along with previous experiences of the scientists on the task force identified some methodology that the task force needed to develop in order to appropriately conduct the simulations and interpret the results. These included detailed instructions on how to select the input parameters and perform the simulations (including keeping experimental results separate from those performing predictive simulations), a program for evaluating the sensitivity of model predictions to input parameters, and statistical procedures for quantifying the fit between model predictions and field data.

5.1 Procedures for Selecting Input Parameters

The task force identified at an early stage (based on personal experience and the literature survey) that providing detailed on selection of input parameters was critical to cold and site-specific comparisons. This is to ensure that the validation focused on the model and the associated procedures rather than the ability of the specific modeler. The task force prepared detailed SOPs for preparing input sequences for both of these comparisons. The intent of these SOPs was to remove all judgment on the part of the modeler. The importance of these SOPs was demonstrated in the initial pilot runs with different modelers. Due to a combination of not specific-enough SOPs and modelers not following the SOPs, very different predictions were obtained. As a result of this experience, changes were made in both the SOPs and the modeling procedures. The task force considers these procedures an important product of the validation research. The SOP for site-specific comparisons is attached as Appendix 2. The guidance for cold simulations presented in Appendix 2 was superseded by a more recent EPA document.

Experimental results were not available to those performing the predictive simulations. One contractor had the job of reviewing the specific data sets and abstracting the relevant input variables as well as the experimental results. The input parameters were then passed onto another contractor who performed the modeling simulations. This two step process minimized bias due to the judgment of the modeler and helped maintain the confidentiality of the chemical(s) used in each study. All test chemicals were identified by a code number rather than by a common name. As a final measure to ensure confidentiality, task force members were not given access to the raw data submitted for each field study.

5.2 Guidance for Performing Calibration Simulations

As mentioned earlier, performing calibration simulations can provide valuable information about selection of input parameters and model performance. However, it is important that calibration simulations do not just consist of a simplistic regression of input parameters to minimize differences between observed and predicted values since most water quality models have enough adjustable parameters to fit a limited set of field observations. In order to make certain that such calibration simulations are performed, a document providing guidance on how these simulations has been prepared and provided to the modelers performing these types of simulations (attached as Appendix 3). Unlike the other comparisons (cold and site-specific), the judgment of the modeler will still have some influence on the performance of these simulations, although it should be minimal if the guidelines are followed.

The guidelines emphasize two principles. The first is that it must be clearly recognized that both the observed field results and the predicted modeling results contain error and neither value should be regarded as absolutely correct. The second is that model input parameters should not be adjusted outside the ranges that are reasonable. If parameter adjustment beyond a reasonable range is required to achieve a satisfactory fit, then there is a potentially a problem with either the model or the experimental data.

The general procedure for calibrating modeling to experimental results (leaching or runoff) was to first calibrate the hydrology of the model to provide a reasonable representation of water movement at the

specific study site. Then, the simulation of the transport and dissipation of the chemical was evaluated and calibrated as necessary.

5.3 Statistics

In the initial phase of evaluating model performance, the FEMVTF statistics committee performed correlative and statistical analysis of measured field data and model predictions paired in time. The unique nature of the correlative exercise of comparing model estimates with measured field data via standard statistical objective analyses, while informative, failed to satisfactorily provide insight into the validity of the model estimates. One of the keys to the success of a model validation study is the collection of high-quality field measurements against which the model predictions will be tested. Standard statistical analysis generally uses paired data, i.e., a model prediction and a field measurement paired in space (e.g., depth) and time. Although, the model response data sets were large, a critical issue encountered was the typically small size of the field data set, particularly for runoff studies. The greater the number of paired data points, the greater the confidence in the model validation study results. For example, a site with less than 5 - 10 of a specific measurement would have reduced power for model testing. The amount of rigorous statistical testing possible for the small data sets was found to be severely limited. For example, for the runoff data sets analyzed, calculating the concordance statistic (I-Kuei Lin, 1989, 1992) with three paired values provided little statistical power. The correlative exercise of comparing model estimates with measured field data via standard statistical objective analyses failed to identify factors associated with the uncertainty of measured environmental fate data and potential model inputs.

Therefore, to evaluate the impact of uncertainty, additional statistical analyses were performed on selected data sets. As an initial step to evaluate the impact of uncertainty, a sensitivity analysis was performed using an approach based on that of Plackett and Burman (1946) to identify key model input parameters for runoff and leaching simulations.

Monte-Carlo analysis was used to evaluate the uncertainty associated with each sensitive input parameter. One of the most important steps in this process was the development of distributions for each key parameter that could be sampled during the Monte-Carlo analysis. Specific criteria were developed for establishing these sampling distributions to ensure consistency in the procedures for evaluating model prediction error across sites and also to ensure that the sampling distributions represented, to the degree possible, the actual site-specific uncertainty and variation in the parameters. Therefore, the criteria effectively increased the confidence that the Monte Carlo uncertainty analysis results reflect the true model prediction error associated with a specific site and parameter set. In addition, the criteria provided a record against which the sampling distributions were judged.

5.4 Sensitivity Analysis

Introduction. Plackett-Burman was chosen as the procedure for conducting sensitivity analyses because of its simplicity and suitability for the identification and ranking of variance components in multiparametric models (Plackett and Burman, 1946). Identification of the limited "primary" variables controlling model outcomes allows for stepwise refinement of the model through identification and control of the "primary" variables. PB analysis isolates the main effect of a variable in a model through a contrast of outcomes at two different levels of the variable. This is accomplished by investigating equal numbers of combinations of each variable at predetermined "high" and "low" levels dictated by the selection of the perturbation factor. The average difference in outcomes over the various combinations of variable input parameters allows for determining the effect of change per unit for each input parameter. Inclusion of a subset of dummy variables produces a model error component that can be used to test the variance of each input parameter of the model.

PB design uses an incomplete block factorial design where each input parameter is evenly weighted. This reduces the number of unique model simulations required at the expense of assuming that

interactions between more than two variables are insignificant, as this variance is grouped with the model error. Designs for contrasts at two levels are presented by Plackett and Burman (1946).

Analysis of the results of a PB design is straightforward. First, the average difference in the sum of "high" and "low" inputs for the test and dummy parameters is determined. The standard error variance of an effect (V) is the variance due to dummy variables (E_{D_i} , where $i = 1, n$ with n being the total number of dummy variables; typically, $n = 5$ degrees of freedom for the PB program):

$$V = \frac{\sum_{i=1}^n (E_{D_i})^2}{n}$$

The standard deviation of the effect is thus,

$$s = V^{1/2}$$

A simple t-test allows for testing of the significance of the effects found for the mean difference between two real variables (X):

$$t = (X_h - X_l)/s_x$$

$$\text{where } s_x = \frac{s}{\sqrt{n}}$$

The result of PB analysis is, therefore, a ranked listing of variables in order of their relative effect on model outcomes along with a determination of relative significance of the effect. PB analysis has been used to evaluate input parameter sensitivity for transport modeling of pesticides at the regional scale (Cryer et al, 1994; Fontaine et al, 1992; Cryer and Havens, 1997).

Soil Correlations to Eliminate Potential Nonsense Parameter Combinations. In addition to the experimental design considerations, several physically based correlations for soil properties are employed to avoid possible "nonsense" parameter combinations that can be obtained from the PB analysis. The following simple equality must be obeyed for all soils.

$$\text{Wilting Point (WP)} \leq \text{Field Capacity (FC)} \leq \text{Porosity (POR)} \quad \{5-1\}$$

If WP, FC, and POR are treated as independent parameters with certain magnitudes, then the PB design could result in violation of equation 5-1. This typically causes the GLEAMS model to generate erroneous results or not run at all. The following soil correlations are implemented for both the GLEAMS and PRZM models to avoid potential nonsense soil parameter combinations.

$$\text{FC}(i) = 0.3486 - 0.0018 \text{ SAND}(i) + 0.0039 \text{ CLAY}(i) + 0.0228 \text{ OM}(i) - 0.0738 \text{ BD}(i) \quad \{5-2\}$$

$$\text{WP}(i) = 0.0854 - 0.0004 \text{ SAND}(i) + 0.0044 \text{ CLAY}(i) + 0.0122 \text{ OM}(i) - 0.0182 \text{ BD}(i) \quad \{5-3\}$$

$$\text{BD}(i) = 2.65 [1 - \text{POR}(i)] \quad \{5-4\}$$

The array subscripts "(i)" in equations 5-2 and 5-3 represent layers i , from $i = 1$ to $i = \text{total number of soil layers that are being modeled}$.

In these correlations, both WP and FC are treated as dependent variables and are functions of soil texture (% CLAY and % SAND), bulk density (BD), and soil organic matter (OM). Thus, % SAND, % CLAY, OM and BD can be investigated in the sensitivity analysis. FC and WP are calculated from these independent variables. Certain parameter combinations substituted into equations 5-2 through 5-4 do not satisfy equation 5-1. In cases where $WP > FC$, WP is set equal to FC minus a small amount. In cases where $FC > PRO$, FC is set to a value slightly less than POR. This guarantees that equation 5-1 will always be true for all independent variable combinations regardless if the parameter combinations were real or imaginary.

Soil Properties which Vary with Depth. Several soil properties can vary with depth throughout the soil horizon. Examples include porosity/bulk density, field capacity, wilting point, organic matter, and pH. The user specifies as input the soil depth increments where property values differ, along with the magnitudes of the properties at each specified depth. The number of PB simulations required if all soil intervals were modeled as being independent would dramatically increase if each soil layer were treated as a separate entity. Therefore, parameters that can change with depth are grouped together and changed according to the original user defined magnitude (via the nominal file) at each depth. For example, if soil organic matter is chosen as an input parameter to investigate, the PB program changes all soil organic matter by the same perturbation factor specified by the user. If the user specifies a 10 percent perturbation around the nominal value and the PB design specifies a "-1", then all of the soil organic matters values for each soil layer and for this simulation are decreased by approximately 10 percent. Thus, the same value for organic matter is not simulated for all soil depth intervals, (unless the user specifies the organic matter does not vary with depth in the nominal file) but rather a consistent and constant percent change occurs for each depth value.

Implementation of Sensitivity Analysis within FEMVTF. Plackett-Burman sensitivity analysis has been an integral part of numerical modeling risk assessment within the DowElanco GRASP and DEGAS systems where it has been used to discern sensitive inputs to chemical transport models (Fontaine et al., 1992; Cryer and Havens, 1993; Cryer et al., 1994; Cryer and Havens, 1997). The sensitivity analysis in DEGAS has been automated and consists of FORTRAN and UNIX shell scripts. The system is robust and usable and fits well with the goals of FEMVTF, therefore, the original code for PB analysis has been re-partitioned and altered by FEMVTF to conform to FORTRAN 90 coding standards and to be hosted in a DOS-based environment. These executables have been linked such that they run in the DOS window on Windows 3.1, '95, or NT as well as from the standard DOS prompt. Interim releases of the PB software (PB 1.0b, PB1.0, and PB 1.01) compatible with interim releases of PRZM3 were used for sensitivity analysis of two leaching and two runoff data sets in the FEMVTF Phase II pilot exercise. Two additional data sets (one run off and one leaching) were analyzed using PB 1.2b3, a version coded for compatibility with the CEAM beta-release of PRZM3 (version 3.12 beta, ftp://ftp.epa.gov/epa_ceam/wwwhtml/softwdos.htm).

PB additionally supports GLEAMS simulations. This aspect of PB is well-documented (Cryer, 1996; Cryer and Havens, 1997) and, therefore, was not considered within FEMVTF.

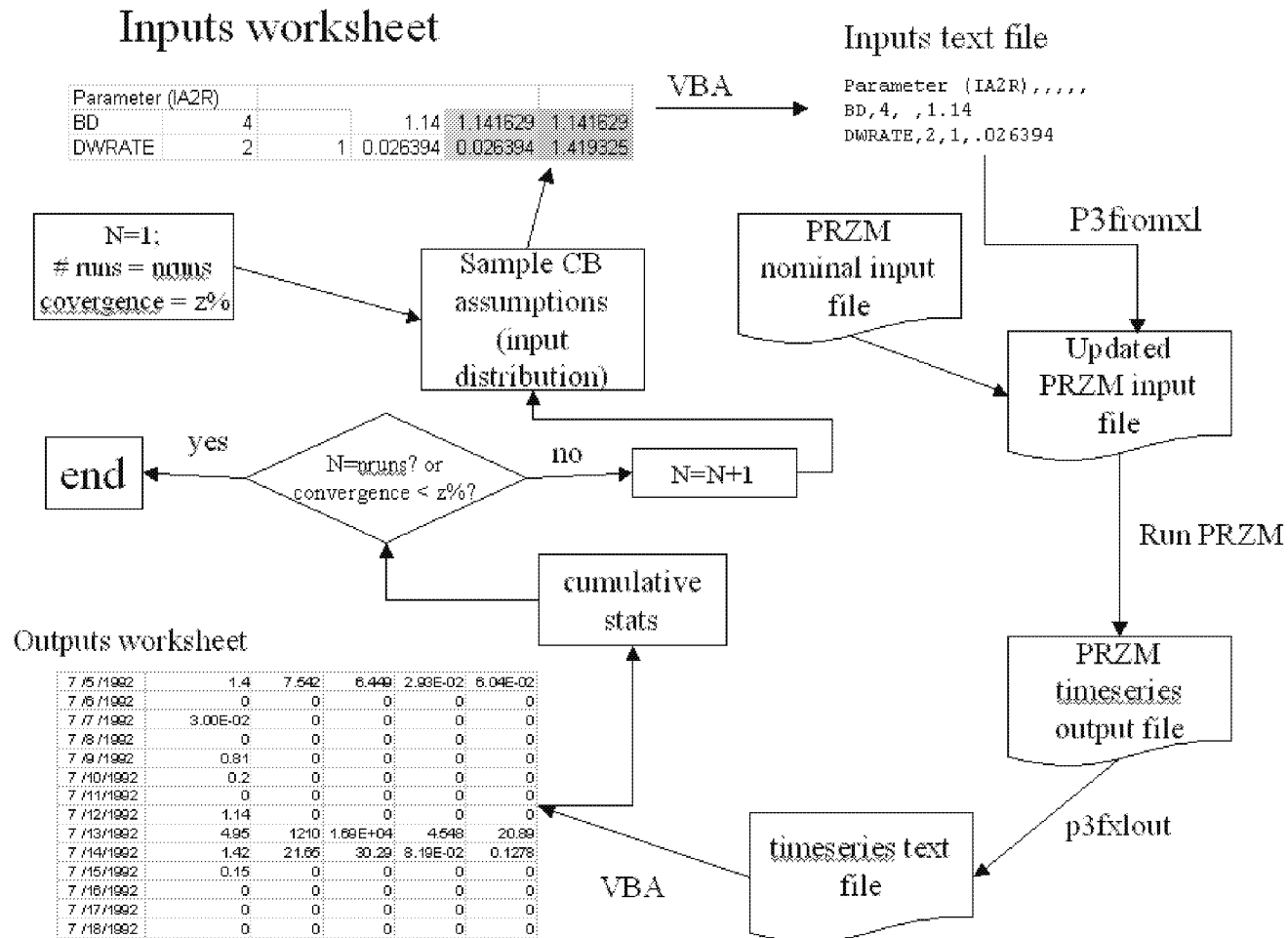
FEMVT Evaluations and Results. Sensitivity analyses were conducted with three leaching and three runoff data input sets were performed in accordance with guidance documentation presented in Appendix 4 for PB version 1.0 (data sets NC1L, NC2L, GA1R, and IA2R) or PB version 1.2b3 (data sets NC3L and GA2R). Results of the PB analysis are summarized in Sections 6 and 7, with the details included in the discussion of the appropriate data sets in Appendices 5 and 6.

5.5 Uncertainty Analysis

Monte Carlo analysis is a powerful tool for conducting uncertainty analysis on complicated models like PRZM3.12. A review of Monte Carlo procedures, a guide to the selection of sampling distributions, and an analysis of proper Monte Carlo procedures is found in Warren-Hicks and Moore (1998). The FEMVTF Statistics Committee paid close attention to the procedural and statistical pitfalls of Monte Carlo analysis. The following activities were implemented as an effort to ensure the correct implementation of

the Monte Carlo analysis: 1) strict guidelines were developed for the selection of sampling distributions for the input parameters (see Appendix 6), 2) numerous information sources, data bases, and experts were identified and consulted in the course of selecting the input parameter sampling distributions, 3) a rigorous evaluation of statistical correlation among the input parameters was undertaken (note: the committee concluded that no statistical correlation exists between the parameters selected for evaluation), and 4) a comprehensive sensitivity testing of the Monte Carlo outputs was implemented in an effort to ensure results that are not overly dependent upon assumptions and interpretations.

Development of Crystal Ball® Monte Carlo Analysis Tool. A set of interface tools was built to implement the Monte Carlo sampling and analysis techniques with the PRZM3.12 model. The software chosen was the Crystal Ball® Pro package, manufactured by Decisioneering, Inc. (Denver, CO), along with some additional FORTRAN programs. Crystal Ball® Pro is an add-on package to Microsoft Excel® (Redlands, WA) which allows the user to define distributions and sampling methods for model inputs as well as store the outputs from multiple-run modeling sets. As much of the functionality of Crystal Ball® Pro is available to Excel®'s Visual Basic for Applications (VBA) programming environment, VBA scripts were developed to set up and manage the Monte Carlo analysis. Figure 5-1 shows the general flow of the interface system.

Figure 5-1. Crystal Ball® Monte Carlo Analysis Tool Flowchart

To set up a simulation set, the user first must define the inputs to be varied and the distribution types describing its behavior. These distributions, termed 'assumptions' by Crystal Ball[®] Pro, were defined as described in the preceding discussion. When a set of runs is initiated via a VBA script, Crystal Ball[®] Pro samples (employing Latin Hypercube methodology) from the pre-defined assumptions to set the input values for the current simulation. The script then writes the input values to a text file. An MS-DOS process is then spawned, executing the p3fromxl FORTRAN program which spawns the text file and inserts the updated input parameter values into an existing PRZM3.12 input file. The PRZM model is then executed with the new file as input. The existing input files were the same files used in the Level 2A analysis with the exception of the USLEC values. The Monte Carlo application was limited to the fallow, cropping and residue values for this parameter. The p3fxlout FORTRAN program parses the resulting PRZM output file and the desired time series outputs are imported back into Excel[®] arrays, termed 'forecasts' by Crystal Ball. The run number counter is then incremented and the loop repeated until the desired number of simulations has been run.

By using the 'keep sorted values' option in Crystal Ball[®] Pro, the package keeps the input and output values in their proper correspondence, so the statistical analyses described above could be done by exporting the output and input forecasts from Excel[®]. In addition, convergence statistics were accumulated for subsets of runs; this showed that about 500 total simulations were sufficient to describe the variability of outputs as a function of the sampled input distributions. The intent of the analysis was not to completely define the prediction uncertainty for any one parameter. Rather, the purpose was to evaluate the relative spread in model predictions around the measured values. Therefore, the number of iterations was not based on precisely estimating the prediction distribution, but to provide a sufficient level of convergence and to bound the expected range of model predictions given the input sampling distributions.

Quality Control/Quality Assurance. A two-tiered quality control check was performed for each parameter of interest such as runoff volume, pesticide mass in runoff, bromide in soil pore-water, parent soil core concentrations at depth, etc. The quality control process was initiated to ensure that for the comparative aspect of the analysis i.e. comparison of the Monte Carlo estimates versus measured data that all data reflect exactly the measured values. Where data were extracted from study reports, analysts reviewed and verified that electronic files created from the study data reflected exactly the study reports. For the second tier, comparisons were made between the electronic field study summaries and the text files used to input the field study data into SAS (SAS Institute, Cary, NC) using Monarch[®] (Datawatch Corp., Wilmington, MA) to ensure that transcription errors would not occur. Where discrepancies attributable to rounding were noted, the impact of the rounding differences on the analysis was assessed and rectified if warranted. Where discrepancies were noted further investigation was initiated and the errors were rectified where warranted. The process ensured that the files used in the SAS comparative analysis did indeed reflect exactly the measured field data.

5.6 References

- Cryer, S. A. 1996. Plackett and Burman Experimental Design Sensitivity Analysis for the Environmental Fate Models GLEAMS, SWRRBWQ, EPICWQ, and PRZM-2. GH-C 3815. Unpublished report of Dow AgroSciences, Indianapolis, IN.
- Cryer, S. A., and P. L. Havens. 1993. The Role of Sensitivity Analysis in Modeling Potential Environmental Concentrations Resulting from Pesticide Laden Agricultural Runoff. DowElanco/Purdue Colloquium, 25—26 May 1993, Purdue University, West Lafayette, IN.
- Cryer, S. A., and P. L. Havens. 1997. Regional Sensitivity Analysis for the USDA Model GLEAMS. Environmental Modeling and Software (submitted).

Cryer, S. A., P. L. Havens, R. A. Swearington, and P. G. Nordstrom. 1994. Addressing Regional Runoff Variability using a Binned Interval Approach. 748 In Abstr. 8th Intrnl. Congr. Pest. Chem., 4—9 July, 1994, Washington, D. C.

Fontaine, D. D., P. L. Havens, G. E. Blau, and P. M. Tillotson. 1992. The Role of Sensitivity Analysis in Groundwater Risk Modeling for Pesticides. *Weed Technol.* 6:716-724.

I-Kuei Lin, L. 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* 45:255-268.

I-Kuei Lin, L. 1992. Assay Validation Using the Concordance Correlation Coefficient. *Biometrics* 48:599-604.

Plackett, R. L. and J. P. Burman. 1946. The Design of Optimum Multifactorial Experiments. *Biometrika* 33:305-325.

Warren-Hicks, W. J. and R. J. Moore, R. J. eds. 1998. *Uncertainty Analysis in Ecological Risk Assessment*. SETAC Press, Pensacola, FL, pp. 276.

6.0 Results of Runoff Simulations

The runoff model validation study was designed with the following objectives:

- To evaluate the performance of PRZM3.12 in predicting runoff volume, sediment loss, pesticide losses with runoff and sediment and pesticide concentrations in runoff and sediment.
- Assess the value of using standard modeling guidelines in comparison with the subjective modeling based on expert analysis.
- Identify limitations of the current modeling procedures and techniques.
- Propose recommendations to refine modeling procedures and techniques that can lead to improved predictions.

A total of nine runoff data sets were processed. These runoff studies represented a wide range of hydrological, agronomical, and physiochemical scenarios. The field area for the runoff study sites varied from 0.06 ha to about 7.0 ha. The slope range was 0.25 percent to about 5 percent. Soil texture varied from silty clay loam to loamy sand. All the runoff study sites were under corn or cotton production except one site, which was under sweet corn production. Also eight of these runoff sites were treated with insecticides and one site with a herbicide. The half-lives of the chemical applied varied from 6 to 165 days and Kd value varied from 0.26 to 3200 ml/g for surface horizons. Table 6-1 describes the main features of the runoff data sets and details are provided in Appendix 5.

Table 6-1. The main features of the runoff data sets.

Data Set	Area (ha)	Slope (%)	Soil Type	Crop	Application Method ¹	Half-Life ² (d)	Kd ² (ml/g)
GA1R	3.64	3.5	Loamy sand	Cotton	Aerial (L)	6	4
GA2R	3.04	3-5.5	Sandy loam	Sweet Corn	Foliar (L)	8	0.43
IA2R	7.0	4.3	Silt loam	Corn	T-Band, foliar, and broadcast (G, L)	30	121
IA3R	0.065	5.6	Silt loam	Corn	T-Band, foliar, and broadcast (G, L)	30	121
IA4R	1.21	2.9	Silt clay loam	Corn	T-Band (G)	52	4200
					T-Band (G)	121	12
IA5R	0.065	2.8	Silt clay loam	Corn	T-Band (G)	52	3200
					T-Band (G)	121	10
KY2R	0.065	4.2-5.2	Silt loam	Corn	T-Band (G) ³	52	2200
					T-Band (G) ⁴	121	6
MD1R	0.50	2.0	Silt loam	Corn	Ground spray (L)	165	0.6
MS1R	2.1	0.25	VF sandy loam	Cotton	Foliar (L)	5.7	4.1

1. Formulation type given in parentheses (G = Granular, L = Liquid)

2. Soil-specific half-life and Kd values for surface horizons reflect the values provided by the registrants.

3. Applied to plot 1 and 2

4. Applied to plot 3

The simulations for model validation study were completed in two phases. The first phase simulations were termed as 'site-specific simulations.' The input data for the site-specific simulations were assembled using FEMVTF guidelines (Level 2 SOP Version 2.1, June 12, 1998), attached as Appendix 2. The modelers were not involved in assembling the input data for site-specific simulations to eliminate any

bias from the modeler. The next phase of the study termed as 'calibration simulations' was designed with an overall objective of minimizing the differences between the observed and predicted values for various output parameters using the guidelines presented in Appendix 3. In general, the calibration simulations involved explorative modeling that ranged from calibration to hypothesis testing to a directed sensitivity analysis (other than that of PB analysis). The calibration simulations were conducted based on specific observations and expert analysis in order to: a) better represent actual study site conditions, 2) examine variability in site properties, 3) address uncertainty in input parameters, or 4) to address inaccuracies in measured data. The input parameters (mainly hydrological and environmental fate) were adjusted based on the additional information obtained from a detailed review of runoff study reports made available to the modeler after the site-specific simulations. The calibration simulations were performed for all runoff sites except for GA2R and MS1R sites. The runoff sites GA2R and MS1R did not undergo the calibration work mainly due to time and budgetary constraints. A detailed description of the analysis for the calibration simulations for each individual data set is provided in Appendix 5.

The major output variables considered for model evaluation included runoff volume, sediment yield, pesticide mass and concentrations in runoff and sediment. For a simple statistical evaluation of the model's performance, ratios of predicted and observed values (predicted value ÷ observed value) were calculated for each output parameter (runoff volume, sediment loss, and pesticide mass and concentrations in the runoff and sediment). These ratios were calculated for individual events as well as for the cumulative values of output parameters (average for concentrations in runoff and sediment) during the study period for each runoff data set.

Alternatively, the scatter plots were prepared for the pooled (all the runoff data sets grouped together) observed and predicted values for a given output parameter (e.g., runoff volume) to evaluate the overall variability between the observed and predicted data. These plots also show the quartile ranking of the events (based on measured data) allowing an assessment of the prediction accuracy as a function of relative magnitude of the event.

Site-Specific Simulations. The predicted vs. observed ratios based on individual events and cumulative values are presented in Table 6-2 and Table 6-3, respectively, for site-specific simulations. Figure 6-1 presents the scatter plots for the site-specific simulations.

Table 6-2. The ranges of ratios (predicted value ÷ observed value) based on individual events for site-specific simulations.

Site	Runoff	Sediment Loss	Pesticide Mass in Runoff	Pesticide Mass in Sediment	Total Pesticide Mass in Runoff and Sediment	Pesticide Conc. in Runoff	Pesticide Conc. in Sediment
GA1R	0.6 to 6.4	0.2 to 1.3	0.7 to 1.2	1		0.3 to 1.3	
GA2R	0.4 to 4.4					0.9 to 9.0	
IA2R	0.8 to 3.8	0.001 to 3.5	0.4 to 29.0	0.007 to 7.2	0.02 to 8.5		
IA3R	0.28 to 1.5	0.08 to 13	0.04 to 3.3	0.04 to 43	0.04 to 7.8	0.14 to 2.6	0.5 to 6.8
IA4R	0.01 to 0.8	0.003 to 0.7	0.0 to 2.1			0.0 to 2.7	
			0.07 to 1.8			2.3 to 5.8	
IA5R	0.01 to 0.8	0.0001 to 11.8	0.0 to 0.89			0.0 to 1.4	
			0.0 to 0.7			0.0 to 8.7	
KY2R ²	0.8 to 2.4	0.4 to 2.3	0.6 to 5.8			0.6 to 2.6	
			2.4 to 6.0			2.8 to 3.1	
MD1R	0.9 to 32		2.3 to 7.4			0.0 to 4.4	
MS1R	1	2.3	0.3			0.3	

1. Empty boxes indicate no measured data available for comparison

2. All data pooled for three subplots.

Table 6-3. The ratios of predicted and observed values (predicted value ÷ observed value) based on cumulative values (average for concentrations) for site-specific simulations.

Site	Runoff	Sediment Loss	Pesticide Mass in Runoff	Pesticide Mass in Sediment	Total Pesticide Mass in Runoff and Sediment	Pesticide Conc. in Runoff	Pesticide Conc. in Sediment
GA1R	1.47	0.28	1.00			0.39	
GA2R	2.03					1.43	
IA2R-1992	1.63	2.56	2.64	1.75	2.03	3.32	3.84
1993	1.19	0.23	3.99	0.32	0.50		
IA3R-1992	1.45	7.96	1.71	3.78	2.66	1.33	1.98
1993	1.02	0.93	0.66	0.51	0.55	0.31	1.20
IA4R	0.34	0.25	0.38			0.83	
			1.01			2.73	
			0.15 ²			0.15 ²	
IA5R	0.57	0.25	0.18			0.27	
			0.40			0.83	
KY2R-plot 1	1.41	1.32	1.00			0.63	
plot 2	1.30	0.78	2.29			1.70	
plot 3	1.24	0.76	4.11			2.90	
MD1R-1990	3.82		3.70			0.41	
1991	4.75		3.67			0.83	
MS1R	1.00	2.34	0.30			0.30	

1. Empty boxes indicate no measured data available for comparison.

2. Value for bromide

The range of ratios given in Table 6-2 demonstrates that the site-specific runoff predictions were generally within one order of magnitude of observed values for all the data sets except for MD1R. The runoff volumes were consistently over predicted for MD1R site. The predicted sediment losses were within one order of magnitude of observed sediment losses for GA1R, IA3R, KY2R, and MS1R data sets. However, sediment losses were under predicted significantly (mainly due to under prediction of runoff) for IA2R, IA4R, and IA5R. The values for pesticide mass in runoff were also roughly within one order of magnitude of observed values, except for IA4R and IA5R data sets. The pesticide mass in runoff was usually under predicted for IA4R and IA5R sites. The measured pesticide mass in the sediment was available only for two runoff sites (IA2R and IA3R) and predicted values for pesticide mass in sediment were within one order of magnitude of observed data except for two events (IA2R day 181 and IA3R day 196). The predicted pesticide concentrations in runoff and sediment were also within one order of magnitude of measured data except for few events.

The predicted vs. observed ratios derived for cumulative values represented a much narrower band around the ideal value of 1.0 in comparison with those derived for individual events. In general, the ranges for predicted vs. observed ratios were 0.3 to 5.0 for runoff volume, 0.2 to 8.0 for sediment loss, 0.2 to 4.0 for pesticide mass in runoff, 0.3 to 4.0 for pesticide mass in sediment, 0.5 to 2.7 for total pesticide mass in runoff and sediment, 0.2 to 3.0 for pesticide concentration in runoff and 1.0 to 4.0 for pesticide concentration in sediment. Thus, model predictions based on cumulative values (over the entire study period) were in better agreement with measured data than those based on individual events.

The plots presented in Figure 6-1 confirm that predicted values were usually within one order of magnitude of measured values (most data-points falling within 10X and 0.1X lines) for all output variables. Figure 6-1 also shows that the variability (degree of scatter around 1X line) between the observed and predicted values of an output variable usually decreased with the increasing magnitude for runoff volume, pesticide mass in runoff, (excluding bromide data), and pesticide concentrations in runoff. In other words, the values for predicted vs. observed ratios (Table 6-2) beyond one order of magnitude are usually associated with very small events. However, this trend was not very clear for sediment loss, pesticide mass in sediment and pesticide concentration in sediment.

Following are some general observations made from the site-specific simulation results:

1. Overall model predictions for site-specific simulations are within one order of magnitude (0.1x to 10x) of measured data when the individual events are analyzed. However, when the cumulative values (values summed over the study period) are compared, usually there is better agreement between the observed and predicted data than that based on individual events.
2. More accurate predictions of runoff and erosion generally lead to more accurate predictions of chemical losses with runoff or sediment, indicating reasonable representations of environmental fate and transport processes in the model.
3. Runoff predictions are generally in better agreements with measurements than sediment predictions. The more scattered sediment predictions (Figure 6-1) indicate greater uncertainty involved in the parameterization of the soil erosion module. For example, the regional rainfall distribution - a non-site specific parameter, used in the erosion submodel for calculating the peak runoff likely does not accurately represent the site specific rainfall intensity. Another source of potential uncertainty associated with erosion prediction is capturing the seasonal variations in the crop cover (C) and Manning's roughness (N) factors.
4. The predictions of the larger rainfall events are generally closer to the measurements than the smaller events (e.g., IA5R). This can be largely attributed to the inherent limitations of the SCS curve number method which is an empirically derived rainfall-runoff model with no consideration for soil water dynamics or rainfall intensity. Also the current parameterization of runoff curve number (a single curve number is used to represent the entire cropping season) does not represent the hydrologic variation associated with various crop growth stages within the 'cropping' period.

Calibrated Simulations. The predicted vs. observed ratios for calibrated simulations based on individual events and cumulative values are presented in Table 6-4 and Table 6-5, respectively. Figure 6-2 presents the scatter plots of observed and predicted values and quartile rankings (as described for Figure 6-1) for calibrated simulations.

Table 6-4. The range of ratios (predicted value ÷ observed value) based on individual events for calibrated simulations.

Site	Runoff	Sediment Loss	Pesticide Mass in Runoff	Pesticide Mass in Sediment	Total Pesticide Mass in Runoff and Sediment	Pesticide Conc. in Runoff	Pesticide Conc. in Sediment
GA1R	0.5 to 4.6	0.4 to 1.7	0.7 to 1.4	¹		0.2 to 1.8	
GA2R							
IA2R	0.08 to 2.9	0.002 to 1.7	0.2 to 23.0	0.04 to 4.1	0.07 to 5.4		
IA3R	0.14 to 1.3	0.05 to 5.7	0.03 to 4.1	0.11 to 23	0.09 to 5.5	0.23 to 4	0.4 to 15
IA4R	0.59 to 0.99	0.3 to 2.57	0.2 to 5.1			0.3 to 5.2	
			0.4 to 1.7			0.8 to 1.7	
IA5R	0.7 to 0.9	0.2 to 10.2	0.08 to 0.89			0.1 to 1.4	
			0.0 to 2.36			0.0 to 3.4	
KY2R ²	0.8 to 1.4	0.5 to 1.2	0.7 to 3.6			0.5 to 2.7	
			1.1 to 2.4			1.3 to 1.8	
MD1R	0.17 to 13		0.3 to 0.86			0.0 to 2.6	
MS1R							

1. Empty boxes indicate no measured data available for comparison

2. All data pooled for three subplots.

Table 6-5. The ratios of predicted and observed values (predicted value ÷ observed value) based on cumulative values (average for concentrations) for calibrated simulations.

Site	Runoff	Sediment Loss	Pesticide Mass in Runoff	Pesticide Mass in Sediment	Total Pesticide Mass in Runoff and Sediment	Pesticide Conc. in Runoff	Pesticide Conc. in Sediment
GA1R	1.21	0.82	1.15	¹		0.54	
IA2R-92	1.36	1.50	2.20	1.16	1.49	3.32	7.08
IA2R-93	0.90	0.32	3.01	0.30	0.43		
IA3R-92	1.21	4.04	1.48	2.47	1.94	1.37	3.76
IA3R-93	0.80	0.84	0.79	1.31	1.16	0.43	2.23
IA4R	0.74	0.53	0.62			1.07	
			0.94			1.27	
			0.39 ²			0.31 ²	
IA5R	0.75	0.44	0.33			0.42	
			0.38			0.50	
KY2R							
plot 1	1.16	1.19	0.85			0.72	
plot 2	1.03	0.60	1.82			1.86	
plot 3	1.00	0.61	1.71			1.61	
MD1R							
1990	1.19		0.80			0.31	
1991	1.68		0.27				

1. Empty boxes indicate no measured data available for comparison.

2. Value for bromide.

The calibrated simulations generally resulted in improved predictions as evident from narrower range of predicted vs. observed ratios calculated for calibrated simulations (Table 6-4) than for site-specific simulations. As in the case of site-specific simulations, the predicted vs. observed ratios derived for cumulative values also indicated a better agreement between predicted and observed data than that based on individual events. Also, the predicted vs. observed ratios based on cumulative values for calibrated simulations indicated a substantial improvement over site-specific results for all output parameters except for average pesticide concentration in runoff and sediment. This was somewhat expected because the calibrated simulations effort focused primarily on improving the predictions for runoff volume, sediment losses, and pesticide masses in runoff and sediment. This procedure did not always result in improved predictions for pesticide concentrations in the runoff and sediment.

Figure 6-2 also shows a reduced variability between the observed and predicted values (less data scatter around 1X line) in comparison with the variability associated with the site-specific predictions (Figure 6-1) for all output parameters.

The fundamental difference between the site-specific and calibrated simulations lies in whether the model parameters are kept independent of experimental data during simulation (such as in site-specific simulations) or are derived from parameter adjustments to better fit the data (calibrated simulations). The conceptual basis of a model is truly tested only when the model can predict a set of measurement data with all parameter values determined independently of the data being predicted. However, since environmental processes are always interconnected and often site-specific, determining parameter values truly independently is difficult. As evident from the site-specific and calibrated simulation results, parameter calibration based on site-specific measured data, and expert analysis is essential even though it may not be sufficient to test the model validity.

A detailed sensitivity analysis was also performed for three runoff data sets (GA1R, IA2R, and GA2R) using the Plackett-Berman sensitivity analysis tool (Appendix 4). The PB analyses indicated that runoff curve numbers, bulk density, partitioning coefficient, and degradation rates were among the most sensitive input parameters affecting pesticide losses in runoff and sediment (Table 6-6).

Table 6-6. Results of Plackett-Berman analyses for runoff expressed as relative importance of sensitive components.

	GA1R (foliar)		IA2R (foliar)		IA2R (soil)		GA2R (foliar)	
	Runoff	Erosion	Runoff	Erosion	Runoff	Erosion	Runoff	Erosion
relative importance of sensitive components								
RO curve number 1							25	26
RO curve number 2	85	71	73	30	37		32	21
Kd (layer 1)		8	5	23	23	63		6
Decay rate (layer 1)	9	4	18	6	15	14		
Decay rate on foliage		7		12			4	4
Bulk density (AWHC)	6		4	14	22	17		
Management factor 2		5		8				4
Plant uptake factor		5		7	3	6		

Monte-Carlo simulations were also performed with IA2R and GA1R to evaluate the effect of uncertainty in the input parameters on the model predictions (Appendix 8). The data set IA2R contained a total of seven runoff events. All measured values of runoff volume fell within the interquartile range (between the 25th and 75th percentile of the prediction distribution) of the model predictions, indicating that the model was very reliable. For sediment yield, measured values fell within the interquartile range for three events, within the bounds of the distribution for two events, and outside the bounds of the distribution for two events. For dissolved pesticide runoff mass, five measured values fell within the interquartile range, and the remaining measurements fell within the bounds of the distribution. For pesticide mass in sediment,

three measured values fell within the interquartile range of the predictions, three fell within the bounds of the predictions, and one fell outside the bounds of the model predictions. At GA1R pesticide runoff mass was the only value measured. For each of the four runoff events for which the measurements were available, the measured value fell within the interquartile range of the model predictions.

Cold Simulations. Cold modeling of the runoff transport was performed for only one data set (GA1R) using PRZM3 (Version 3.12, April 21, 1998) and the guidelines prepared for cold simulations (Parker, R., PRZM Inputs - Level One FEMVTF Validation. Revision 11 May, 1999). The predicted vs. observed ratios for cold simulations based on individual events and cumulative values (for runoff volume, sediment loss, pesticide mass in runoff, and pesticide concentrations in runoff) are presented in Table 6-7.

Table 6-7. The ranges of ratios (predicted value ÷ observed value) based on individual events and cumulative values for cold simulations performed with GA1R (pesticide concentrations in sediment were not measured in this data set).

Comparison	Runoff	Sediment Loss	Pesticide Mass in Runoff	Pesticide Conc. in Runoff
Individual Events	0.5 to 2.8	0.05 to 0.3	0.4 to 120.0	0.4 to 42.0
Cumulative Values	1.0	0.1	3.3	1.6

Comparisons of cold simulation results and observed values for runoff volume indicated that predicted results were within one order of magnitude of observed values. However, the sediment yield was consistently under predicted indicating a need for refined/modified erosion parameters. The pesticide mass and concentrations in runoff were substantially over predicted (e.g., 3 to 120 times). The over estimations in the cold simulations were likely caused by the highly conservative estimates of foliar and soil degradation rates for the simulated chemical.

Thus, the general observations from the cold simulations are: a) the cold simulations provide reasonable estimates of the runoff volume; b) the results for sediment yield and chemical mass and concentration in runoff indicate a need for improved parameterization of erosion and chemical environmental-fate parameters.

Conclusions. The overall conclusions drawn from the site-specific and calibrated simulations are:

- The overall model predictions for individual events are usually within one order of magnitude of measured data. When the cumulative or average values (values summed or averaged over the study period) are compared, the agreement between the simulated and measured values is improved. For example, predicted pesticide concentrations in the runoff (averaged over the study period) are approximately within a factor of 0.3x-3x of measured values for both site-specific and calibrated simulations.
- More accurate predictions of runoff and erosion generally lead to more accurate predictions of chemical losses with runoff and sediment indicating reasonable representations of environmental fate and transport processes in the model.
- The variability between the predicted and measured values decreases with the increase in magnitude or event size. For example, the variability of more than 'one order of magnitude' is usually associated with very small events.

Recommendations. Several recommendations are proposed below that would improve the agreement between the simulations and the measurements. However, some of these recommendations may have little relevance in a standard regulatory scenario in which input parameters are usually fixed and event by event match between simulated and measured results is not the objective.

- (1) Although observed and predicted hydrological balances could not be compared (mainly due to a lack of comprehensive measured data), there is a general concern about the model's ability to adequately represent evapotranspiration (ET). ET can affect soil moisture, which can in turn affect the relative daily CN on a runoff day. To improve the predictions for runoff volumes (and sediment losses) hydrologic balance calculations should be considered. Soil moisture and bulk density may also impact ET, and their significance in ET calculations may need to be investigated. In addition, the model needs to account for upward water movement because of the ET in the upper soil profile. The ET extraction depth in the topsoil is somewhat arbitrary and upward water movement is not considered in PRZM3. Evaluating the effects of soil water content at field capacity and soil water content at the wilting point on ET and runoff volume would be useful in better representing the soil water dynamics and overall hydrological balance.
- (2) Although some variability is expected between observed and predicted soil loss values due to empirical nature of soil loss equations, the predictions may be improved by a better representation of storm intensity in the soil erosion submodel. Currently the peak runoff rates in the erosion model are derived from generalized regional rainfall distributions. A better representation of the rainfall distribution may be helpful in improving the soil loss predictions for individual events.
- (3) The latest PRZM-3 (Version 3.12) allows multiple sets of input values for crop cover (C) and Manning's surface roughness coefficients (N). A more detailed description of C and N factors during the cropping period represents the dynamic nature of crop cover and roughness and improves the sediment loss predictions.
- (4) A seasonal variation in runoff curve numbers (similar to C and N factors) may be helpful in representing the effects of changing crop growth stages on predicted runoff. Also, further investigations are warranted for determining the source of discrepancies and improving the model predictions for smaller runoff events.
- (5) The actual time and extent of maximum canopy coverage may vary depending on how well the crop is growing. The extent of maximum canopy and time of maximum canopy, in turn affects the interception and therefore pesticide losses with runoff and sediments. The time and extent of maximum canopy cover calculated from measured canopy cover data can improve model predictions for interception and washoff. The maturation date in PRZM input sequence should represent the time of reaching maximum canopy cover for a given crop.
- (6) A large amount of uncertainty is also associated with the physiochemical properties. The selection of these properties by the registrant remains subjective. A standard procedure may need to be developed for determining the physiochemical properties for the modeling purposes. The following steps may help improving the fate and transport predictions: a) investigating and representing the effects of time and temperature on half-life and Koc on chemical fate and transport, and b) investigating how well the lab values can be extrapolated to the field situation.
- (7) The non-uniform extraction model currently used in PRZM3 does not account for seasonal variations in soil condition and texture. For example, a freshly tilled porous soil would have different pesticide and extraction characteristics than a compacted soil. There can be a future option in PRZM3 to allow the extraction curve to vary by site or over time.
- (8) Site specific situations (e.g., a runoff event spanning over a period of multiple days) need to be carefully represented in the simulation by adjusting the available input/output parameters. Also, the

environmental fate parameters (e.g. half-life and K_{oc}) need to be carefully selected for specially formulated chemicals to represent a realistic environmental fate and transport of these chemicals.

- (9) The sampling inaccuracies should be carefully noted when analyzing the discrepancies between the measured and predicted results. For example, the study report for the IA2R site indicated termination of sampling due to inundation of a primary sampling flume during the runoff event on JD 185-186 in 1993 which would make the observed value suspect for this date.

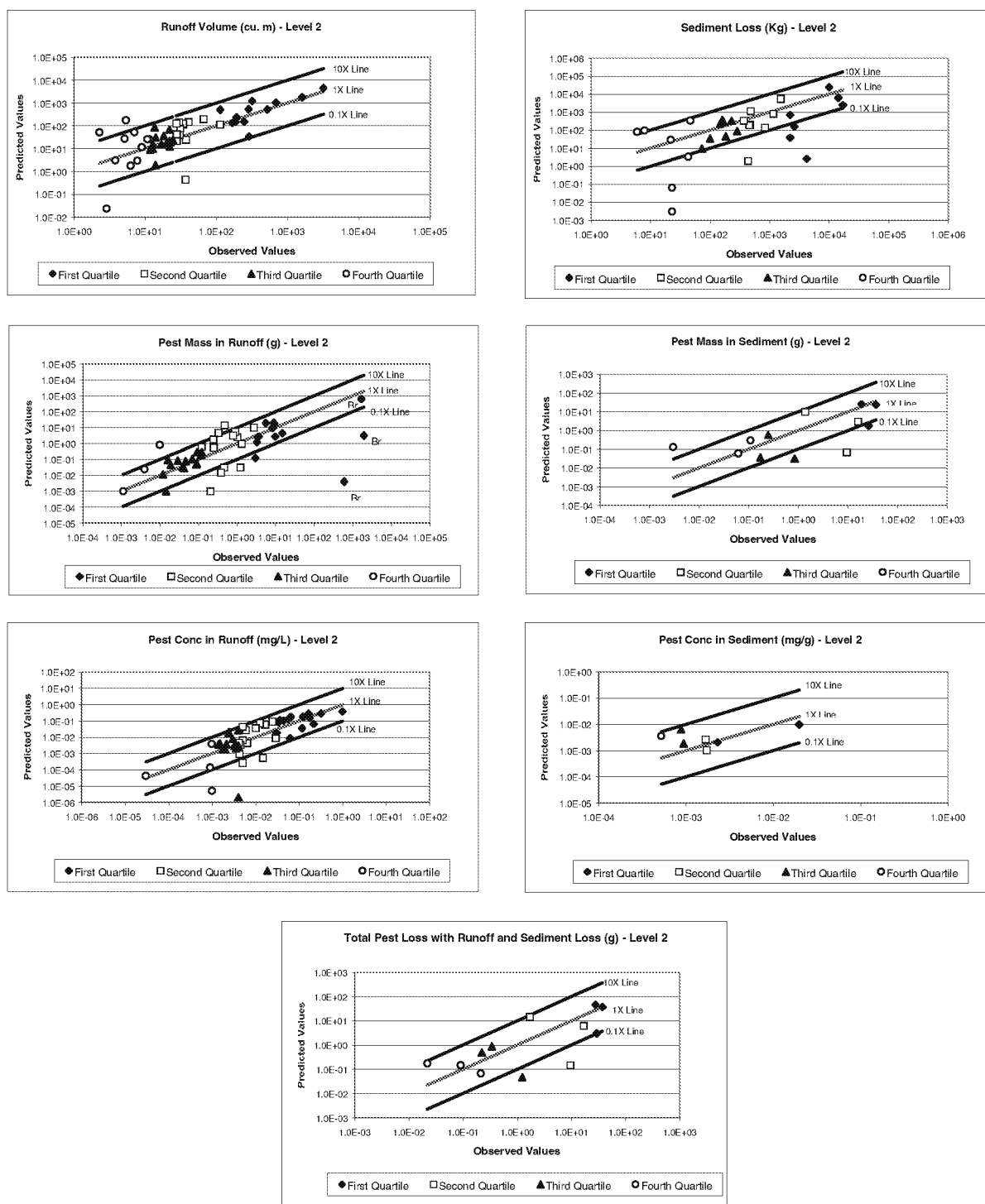
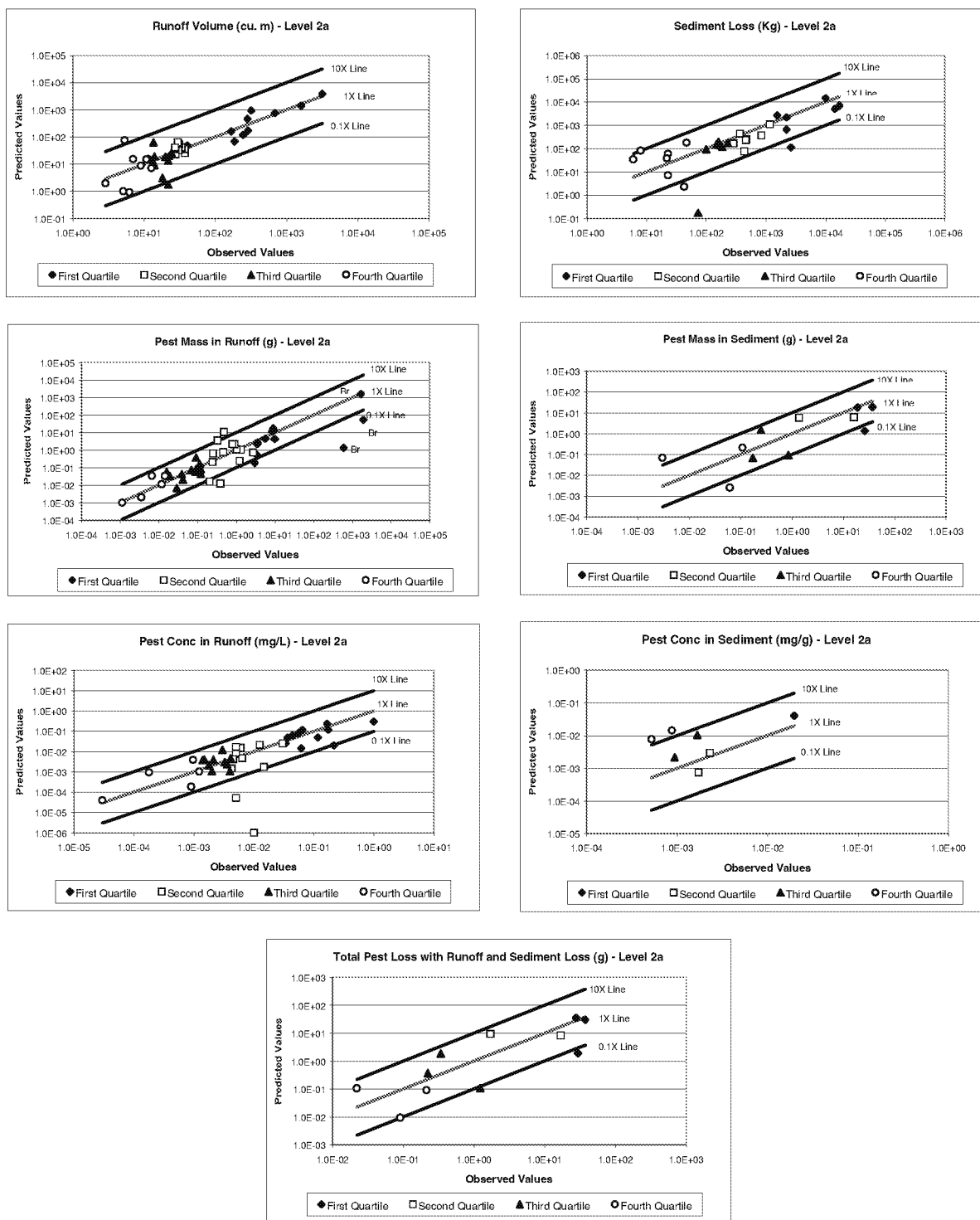
Figure 6-1. The scatter plots of predicted and observed values for site-specific simulations.

Figure 6-2. The scatter plots of predicted and observed values for calibrated simulations.

7.0 Results of Leaching Simulations

The capability of PRZM 3.12 to simulate the leaching behavior of agricultural chemicals was evaluated with the following objectives:

- To determine the performance of PRZM 3.12 in predicting the rate of ground water recharge, based on the comparison of the fate and transport of a non-adsorbed tracer (bromide ion) through the unsaturated zone with measured concentrations in soil and soil pore-water.
- To determine the performance of PRZM 3.12 in predicting the fate and transport of a range of pesticides through the unsaturated zone by comparison with measured concentrations in soil and soil pore-water.
- To assess the relative results of performing leaching assessments using three types of input values: regulatory or “worst-case” (Level 1) input values, standardized input values based on expert judgment and site-specific information (Level 2) and calibrated input values (Level 2a).
- To identify limitations of current modeling procedures and algorithms.
- To propose recommendations for the refinement of modeling procedures and algorithms in order to provide improved prediction capabilities.

A total of nine ground water studies were evaluated. These studies represent a wide range of hydrologic, edaphic and agronomic conditions in agricultural settings located in California, Florida, Georgia, Kansas, and North Carolina. The soil texture at these sites was predominately hydrologic group A soils (typically sand or loamy sand, with a minimum infiltration rate of 8-11 mm/hr). Two sites had less vulnerable loamy sand or sandy loam soils that were classified as hydrologic group B/C or C.

A wide range of crops was grown on the test plots including alfalfa, citrus, sweet corn, peanuts, wheat, soybeans and cotton. The soil degradation half-life of the agricultural chemicals ranged from 8 to 365 days and the sorption coefficient (K_d) varied from 0.088 to 0.50 ml/g for the surface horizon. Table 7-1 describes the main features of the leaching data sets.

Table 7-1. The main features of the leaching data sets.

Data Set	Soil Association	Soil Type	Hydrologic Group	Crop	Application Rate (kg ai/ha)	Half-Life ¹ (d)	K_d ¹ (ml/g)
CA1L	Hilmar	Loamy sand	B	Alfalfa	0.848	115	0.186
FL1L	Astatula	Sand	A	Citrus	5.6	17.5	0.088
GA1L	Kershaw	Fine sandy loam	A	Sweet corn	12.6	8	0.176
GA2L	Ardilla, Clarendon	Mixture, ~sandy loam	B / C	Peanuts	2.49 2.35	69 20	0.42 0.11
KS1L	Las Animas	Sandy loam	C	Wheat	0.074	21	0.50
NC1L	Kenansville	Loamy sand	A	Soybean	0.56	100	0.361
NC2L	Tarboro	Loamy sand	A	Soybean	0.14	47	0.425
NC3L	Tarboro	Loamy sand	A	Cotton	0.14	14	0.091
NC4L	Tarboro	Loamy sand	A	Soybean	0.56	365	0.35

1. Site-specific half-life and K_d values for surface horizons reflect the values provided by the registrants.

Three types of simulations were performed for the model validation study. The first type of simulation was termed a cold (Level 1) simulation and represented the type of conservative evaluation performed by the USEPA for regulatory evaluation of the leaching potential of a pesticide. This predictive approach followed a SOP developed by the USEPA/EFED personnel (Parker, R., PRZM Inputs - Level One FEMVTF Validation. Revision 11 May, 1999) and used extreme input values for some of the input

parameters to generate conservative (sometimes called “worst-case”) estimates of environmental concentrations due to leaching. This approach was expected to generate simulated values as high or higher than the experimental results obtained in prospective ground water studies conducted in highly vulnerable settings. Since only a single “worst-case” combination of input values is simulated in this approach, the probability of actually observing the simulated concentrations is not known.

Due to a combination of logistical problems in completing the cold modeling SOP as well as budgetary and time constraints, cold modeling was only performed for NC4L. This data set was selected due to the fact that it contained the highest ground water concentrations of the nine data sets being evaluated and thus provided the best opportunity to compare the results of progressively refining the accuracy of modeling through the three levels.

Site-specific simulations were intended to represent the results of modeling each ground water study by using standardized “expert judgment” concerning selection of chemical, soil and agronomic parameters. The input data for site-specific simulations were assembled using guidance developed by FEMVTF for uncalibrated or “expert judgment” modeling (Level 2 SOP, Version 2.1, June 12, 1998), attached as Appendix 2. The only study-specific parameters that were included in the site-specific modeling were the rate(s) and date(s) of chemical application and the daily climatic data which were collected at each study site. The resulting site-specific input data sets were then run and post-processed to provide assessment of the hydrology based on bromide fate and transport as well as the fate and transport of the pesticide in soil and soil pore-water at various depths.

The final type of modeling was termed as ‘calibrated simulations’ (Level 2a) and was designed with the overall objective of minimizing the differences between the observed and predicted values by varying key input parameters within experimentally observed ranges. Calibrated modeling work involved exploratory modeling that ranged from calibration of key parameters to sensitivity analysis using either simple variation of input parameters (Appendix 3) or a formal tool such as Plackett-Burman analysis (Appendix 4).

The site-specific simulations were conducted based on specific observations and expert analysis in order to:

- better represent actual study site conditions
- examine variability in site properties
- address uncertainty in input parameters
- address inaccuracies in measured data

Key input parameters (mainly hydrological and environmental fate) were adjusted based on the additional information obtained from a detailed review of leaching study reports made available to the modeler after the site-specific simulations were completed. Calibrated simulations or sensitivity analyses were performed for all leaching studies except for two. The leaching sites CA1L and GA2L did not undergo calibrated modeling or sensitivity assessment primarily due to either time and budgetary constraints or a low benefit from improving the fit obtained in site-specific modeling. A detailed description of the calibrated modeling analysis for each study is given in the respective leaching chapters.

For each level of modeling, the major output variables considered for model evaluation included hydrology parameters (precipitation, evapotranspiration, runoff, erosion and recharge), pesticide fluxes and transformation rates (runoff, erosion, volatilization, plant uptake, foliar dissipation, degradation and leaching) as well as concentrations of bromide and pesticide in soil and soil pore-water over time. For a simple statistical evaluation of the model's performance, ratios of simulated and experimental values (simulated value / experimental value) were calculated for concentration values in soil and soil pore-water over time. These ratios were calculated for both bromide and pesticide concentrations for depths and time points for which experimental data were reported.

One of the obstacles in comparing simulated and experimental leaching data arises from the fact that while modeling can provide both concentration data (mass per volume) and flux data (mass passing through a surface or plane over time), field studies typically provide only concentration data. In addition, annual-average concentration values can readily be calculated from the daily values produced by modeling but time-averaged values are less accurately determined from the monthly sampling events used in most ground water studies.

For regulatory purposes, the primary endpoints are the peak and annual average concentrations of pesticide in ground water. The rate of travel (time of appearance) of peaks is generally of secondary importance. To provide simple but meaningful comparisons between the simulated results and the experimental data, peak concentrations predicted by the model and observed in field monitoring were compared for four endpoints: bromide in soil, bromide in soil pore-water, pesticide in soil and pesticide in soil pore-water. Details of these comparisons for each sampling event are provided in the individual chapters for each study site.

A summary of the mass balances obtained in each of the leaching simulations is provided in Table 7-2. Due to lack of degradation and volatility, the only two dissipation processes for bromide were plant uptake and leaching. For all of the leaching simulations, the PRZM plant uptake parameter (UPTKF) was set at 1.0, the default value recommended in the PRZM manual, which implies that the chemical freely enters plant roots along with transpired water. The resulting plant uptake of bromide that was simulated varied dramatically, ranging from 2-7% in three settings to more than 50% in three settings. For poorly sorptive compounds such as those included in this assessment, the primary factors which influence the plant uptake are the rooting depth (AMXDR) and the recharge that occurs in the months immediately following application. Deeper rooted crops such as alfalfa and citrus can abstract water and bromide from a greater depth than more shallowly rooted crops. In addition, more arid climates, such as California and Kansas, have lower rates of ground water recharge following application which keeps the bromide in the root zone for a longer period of time than in settings with more spring and summer recharge. Based on numerous published studies, typical uptake of bromide into agronomic crops ranges can easily range up to 50% or more which supports the results obtained in this exercise.

Table 7-2. Summary of mass balances for bromide in leaching simulations.

<i>Mechanism</i>	CA1L	FL1L	GA1L	GA2L	KS1L	NC1L	NC2L	NC3L	NC4L
Drift	1.00	1.00	1.00	1.00	1.00	ND	1.00	1.00	1.00
Runoff	0.00	0.03	0.03	0.10	1.31	ND	0.00	0.00	0.00
Erosion	0.00	0.00	0.00	0.00	0.00	ND	0.00	0.00	0.00
Plant Uptake	76.20	58.05	16.78	1.41	52.53	ND	2.78	48.07	6.76
Degradation in 1 m of soil	0.00	0.00	0.00	0.00	0.07	ND	0.00	0.03	0.05
Foliar dissipation	0.00	0.00	0.00	0.00	0.00	ND	0.00	0.00	0.00
Volatilization	0.00	0.00	0.00	0.00	0.00	ND	0.00	0.00	0.00
Leaching below 1 m	22.81	40.92	82.19	97.49	45.09	ND	96.22	50.91	92.19
Remaining in soil	0.00	0.00	0.00	0.00	0.00	ND	0.00	0.00	0.00

ND = not determined

An overview of the comparisons for the peak bromide concentrations in soil and soil pore-water are presented in Tables 7-3 and 7-4. In general, site-specific modeling using PRZM 3.12 provides a very reasonable fit to both soil and soil pore-water data for bromide. The experimental soil data was typically obtained from the soil surface to a maximum depth of 120 cm. The simulated bromide concentrations in soil generally agree with the experimental data within a factor of 3X (i.e. the ratios of simulated to experimental data range from 0.33 to 3.0). Many of the fits were within a factor of 2X. The only notably poor fit resulted at deeper soil depths in NC3L where the experimental concentrations of bromide in soil declined rapidly to very levels less than 0.05 ppm at depths of 90 to 120 cm. The simulated bromide concentrations also declined with depth but not as rapidly as the experimental data.

Table 7-3. The ranges of bromide soil concentration ratios (simulated value / experimental value) for leaching simulations.

Site	Ratio of Peak Soil Concentrations (simulated / experimental)	Depth Interval for Data (cm)	Modeling Level
CA1L	0.19 – 0.74	0 – 60	site-specific
FL1L	0.12 – 3.05	0 – 300	site-specific
GA1L	0.47 – 1.11	0 – 120	site-specific
GA2L	0.30 – 1.11	0 – 120	site-specific
KS1L			site-specific
NC1L			site-specific
NC2L			site-specific
NC3L	0.4 – 73.71	0 – 120	site-specific
NC4L	0.47 – 3.24	0 – 120	cold
	0.40 – 2.71	0 – 120	site-specific

Table 7-4. The ratios of bromide soil pore-water concentrations (simulated value / experimental value) for leaching simulations.

Site	Ratio of Peak Soil Pore-Water Concentrations (simulated / experimental)	Depth Interval for Data (cm)	Modeling Level
CA1L	0.01 – 0.12	90 – 360	site-specific
FL1L			
GA1L	0.74 – 1.26	90 – 360	site-specific
GA2L	1.96 – 3.92	150 – 270	site-specific
KS1L	0.39 – 1.77	90 – 198	site-specific
NC1L			site-specific
NC2L	0.77 – 1.82	90 – 360	site-specific
NC3L	0.33 – 1.66	90 – 360	site-specific
NC4L	1.38 – 3.11	90 – 210	cold
	1.59 – 3.36	90 – 210	site-specific

Blank box indicates no data available for comparison

The simulated concentrations of bromide in soil pore-water also showed excellent agreement with measured data with general agreement typically within a factor of 3X. The measured soil pore-water data was obtained from suction lysimeters installed at soil depths ranging from 90 to 360 cm so this data corresponds to a longer time of travel than the soil data. The slightly different input values determined for NC4L in the cold and site-specific modeling produced slightly different simulated bromide concentrations with slightly better agreement in soil concentrations but slightly worse agreement in bromide soil pore-water. Overall, the capability of PRZM 3.12 of simulating the fate and transport of bromide in the nine ground water studies was very reasonable.

One of the values of using bromide in the ground water studies is to obtain information on the rate of ground water recharge. The simulated recharge rates for the nine studies varied from a low of 29% of applied rainfall and irrigation in California and Kansas to a high of 59% in North Carolina. The mean simulated recharge rate was 45% of applied water. These values agree well with expected recharge rates for shallow ground water in vulnerable agronomic settings.

The simulated mass balances for the pesticides in the nine leaching studies are summarized in Table 7-5. Based on this summary, the major dissipation mechanisms for the various pesticides were degradation and plant uptake with these two mechanisms totaling 58% to 98% of all dissipation (average of 88%). Similar to the plant uptake of bromide, the uptake of pesticides was also significant, ranging from 7 to 51% of the applied chemical. Plant uptake is expected to be highest for pesticides with moderate to long half-life values which are weakly sorbed to soil, a description which fits most of the test chemicals in the cited leaching studies. However, this uptake is higher than would be expected on the existing limited data, probably due to the simplicity of the uptake model used in PRZM. Foliar dissipation of the pesticide was significant for the pesticide in study GA1L, accounting for 39% of the overall dissipation of the chemical.

The simulated runoff and erosion losses were minimal for the leaching studies which agrees well with the fact that these studies were typically sited on flat sites on highly permeable hydrologic group A soils. The two sites with runoff of approximately 0.6% were conducted on hydrologic group B or C soils.

The fraction of applied pesticide that was simulated to leach below a depth of 1 m varied between 0% (KS1L) and 36% (NC4L) with a mean value of 8%. The studies with the highest leaching percentages (CA1L, NC1L and NC4L) all had application rates of more than 0.5 kg ai/ha, sorption coefficients of 0.36 g/ml or less and soil half-lives of 100 days or more. The screening model SCI-GROW suggests that the ground water detections resulting from these combinations of use rate, sorption and half-lives would be

expected to result in ground water detections of 3 ug/L or more which agrees with the results obtained from the deepest lysimeters in these studies.

Table 7-5. Summary of mass balances for pesticide in leaching simulations.

<i>Parameter</i>	CA1L	FL1L	GA1L	GA2L		KS1L	NC1L	NC2L	NC3L	NC4L
Drift	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Runoff	0.00	0.03	0.16	0.59	0.58	1.08	0.00	0.10	0.00	0.00
Erosion	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Plant Uptake	50.80	40.64	6.95	33.05	28.86	30.03	46.28	23.20	31.08	48.47
Degradation in 1 m of soil	36.61	55.02*	50.67	62.65	68.87	67.89	36.11	71.77	64.31	14.86
Foliar dissipation	0.00	0.00	38.96	0.00	0.00	0.00	0.26	0.00	0.00	0.00
Volatilization	0.00	0.00	1.52	0.03	0.00	0.00	0.00	0.02	0.00	0.00
Leaching below 1 m	11.59	3.31*	0.74	0.73	0.58	0.00	16.35	3.90	3.60	35.67
Remaining in soil	0.00	0.00	0.00	1.95	0.11	0.00	0.00	0.00	0.00	0.00

* in top 3.6 m of soil

A summary of the ratios of simulated to experimental results for peak pesticide concentrations in soil is presented in Table 7-6. The simulated soil concentrations of pesticide varied between extreme under prediction (e.g. ratio near 0) to up to 86X over prediction.

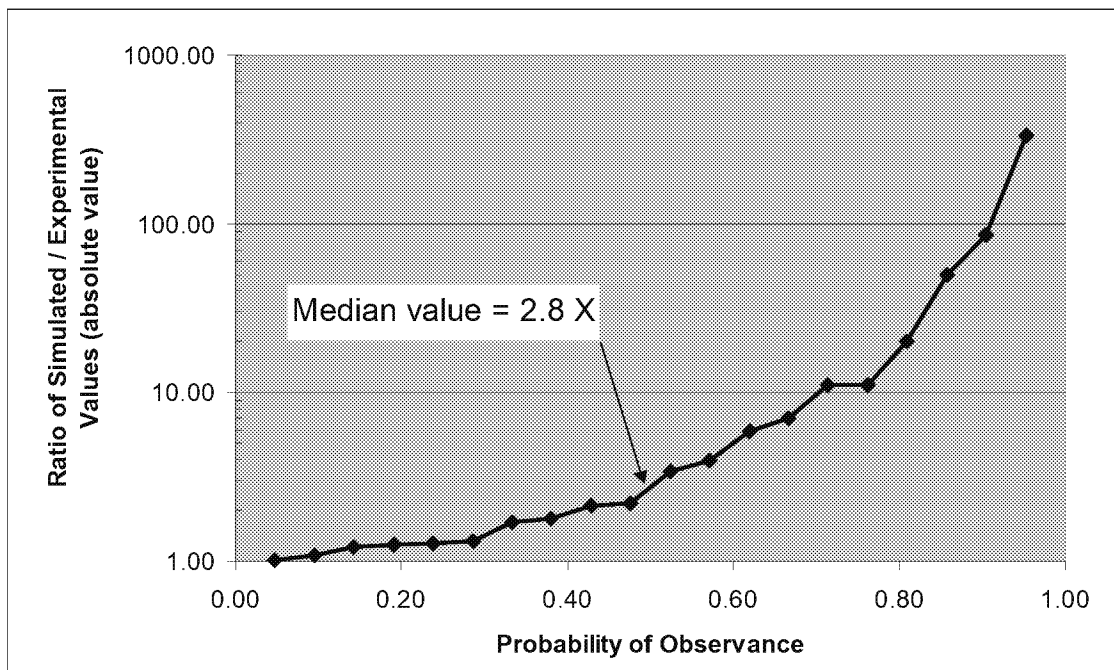
Table 7-6. The ranges of pesticide soil concentration ratios (simulated value / experimental value) for leaching simulations.

Site	Ratio of Peak Soil Concentrations (simulated / experimental)	Depth Interval for Data (cm)	Modeling Level
CA1L	0.56 – 1.21	0 – 60	site-specific
FL1L	0.05 – 0.17	30 – 300	site-specific
GA1L	1.09 – 85.6	0 – 120	site-specific
GA2L	0.09 – 1.70	0 – 120	site-specific
	0.02 – 1.02	15 – 120	site-specific
KS1L	0.003 – 0.78	0 – 90	site-specific
NC1L	0.8 – < 3.39	0 – 220	site-specific
NC2L	0.09 – 0.76	0 – 120	site-specific
NC3L	0.45 - < 3.93	0 – 120	site-specific
NC4L	0.49 – 12.04	0 - 120	cold
	0.47 – 7.08	0 - 120	site-specific

A graphical plot of the ratios listed in Table 7-6 is presented in Figure 7-1. In this figure the absolute values of the lowest and highest ratios from each study have been rank ordered and analyzed to provide an estimate of the median expected agreement. For pesticide concentrations in soil, the simulated

values are expected to be within 3X of experimental values up to 50% of the time. 80% of the time the values are expected to be within a factor of 10-11X. Less than 10% of the time the simulated results are expected to be relatively poor with ratios of 100X or more.

Figure 7-1. Accuracy of simulated pesticide concentrations in soil (since this is a plot of the absolute value of the log of the simulated/experimental data, the accuracy of the simulation is evaluated without regard to whether the result was an over-prediction or under-prediction).



The simulated concentrations of pesticide in soil pore-water also varied widely, with a clear bias toward under prediction (Table 7-7). The highest over prediction was a factor of 16X for NC3L while four studies had significant under prediction of pesticide residues in soil pore-water.

To estimate the decline in degradation rate with depth in the soil profile, the site-specific SOP specified that the soil profile be divided in thirds with degradation half-lives of 1X, 2X and 3X, respectively with depth. This approximation reflects the general recognition that the rate of degradation declines with depth and uses a simple approach to estimate this effect. In several of the ground water studies, this simulation approach generated more degradation than was experimentally observed and resulted in under prediction. For compounds with significant leaching concerns, it may be necessary to obtain one or more experimental measurements of degradation rate in subsoils to ensure reasonable accuracy in simulating the leaching of trace concentrations of pesticides from the root zone. In the absence of experimental rate data, a more rapid rate of decline in degradation rate with depth would help eliminate the under prediction problem.

Table 7-7. The ratios of pesticide soil pore-water concentrations (simulated value / experimental value) for leaching simulations.

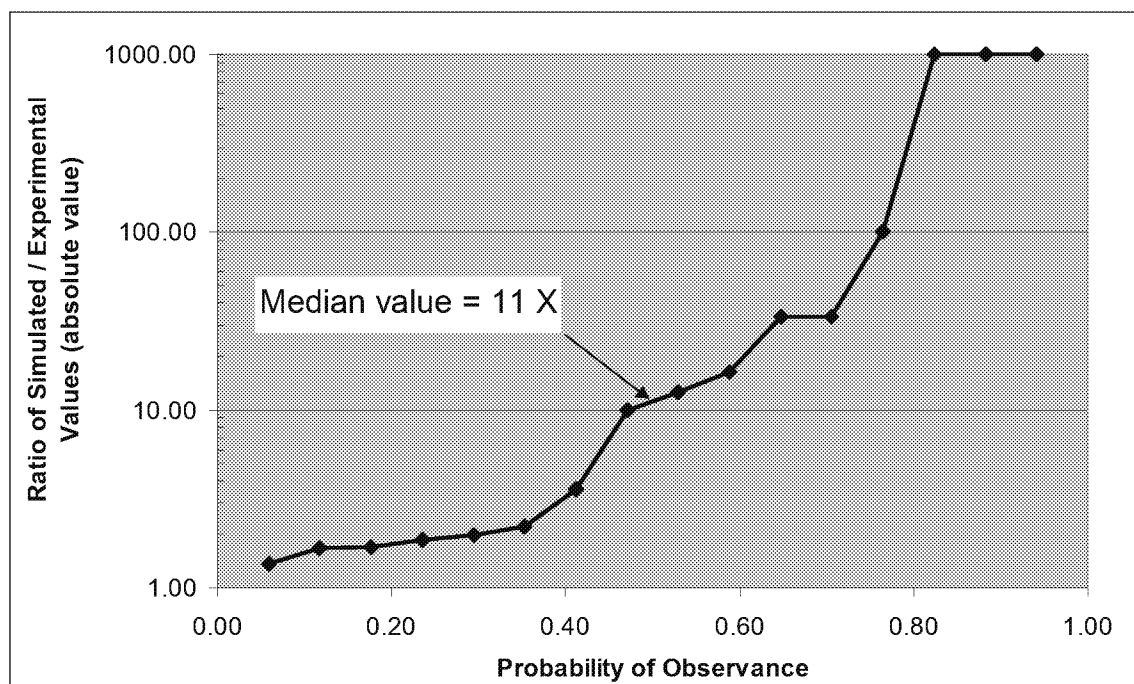
Site	Ratio of Peak Soil Pore-water Concentrations (simulated / experimental)	Depth Interval for Data (cm)	Modeling Level
CA1L	0.01 – 0.10	90 – 360	site-specific
FL1L			
GA1L	0.03 – 1.7 0.43 – 0.90	180 – 360 180 – 360	site-specific calibrated
GA2L	0.03 – 0.08 (Poor fit, see Appendix 6)	150 – 270	site-specific
KS1L	0.00 (Poor fit, see Appendix 6)	90 – 198	site-specific
NC1L	0.76 – 1.36	90 – 270	site-specific
NC2L	0.28 – 1.68	90 – 360	site-specific
NC3L	0.54 – 16.45	90 – 360	site-specific
NC4L	3.53 – 4.01 1.99 – 2.22	90 – 210 90 – 210	cold site-specific

Blank box indicates no data available for comparison

The absolute values of the ratios of peak soil pore-water concentrations for the highest and lowest values from each study were plotted in Figure 7-2. Based on this figure, the median expected accuracy of simulating pesticide concentrations in soil pore-water following the site-specific SOP was approximately 11X. The simulation was within a factor of 100X up to 80% of the time.

The prediction of trace-level residues in soil pore-water or ground water is a technically difficult challenge, especially for multiple applications of pesticides at relatively high use rates. However, the results obtained in this modeling exercise indicate that reasonably accurate values can be obtained in many cases. There were a number of studies that resulted in ratios of approximately 2X which is excellent agreement for a model such as PRZM 3.12. For NC4L, both cold and site-specific modeling were performed. The cold modeling agreed with the experimental data within approximately 4X while the site-specific modeling improved this agreement to within a factor of 2X. This good agreement in this case resulted from the fact that the pesticide degraded rather slowly in the soil profile and the adjustment of the half-life with depth did not significantly affect the predicted concentrations.

Figure 7-2. Accuracy of simulated pesticide concentrations in soil pore-water (since this is a plot of the absolute value of the log of the simulated/experimental data, the accuracy of the simulation is evaluated without regard to whether the result was an over-prediction or under-prediction).



The overall agreement between the simulations and the experimental data have been summarized in Table 7-8. Following the SOP for site-specific modeling, bromide concentrations in soil and soil pore-water can typically be simulated within a factor of 2-3X of experimental results. Pesticide concentrations can be simulated within a factor of 3X in soil in the top 120 cm of the soil profile. Pesticide soil pore-water concentrations can typically be simulated within a factor of 11X to depths of up to 360 cm.

Table 7-8. Summary of median ratios of simulated/experimental values for bromide and pesticide from leaching simulations.

Parameter	Depth Intervals (cm)	Median Ratio (Typical Expected Accuracy)		
		Cold	Site-Specific	Calibrated
Bromide Concentration	0 – 120	3 X	2 – 3 X	ND
	100 – 300	3 X	2 X	ND
Pesticide Concentration	0 – 120	4 X	3 X	ND
	100 – 300	typically > 11 X	11 X	2 - 3 X

A detailed sensitivity analysis was also performed for three runoff data sets (NC1L, NC2L, and NC3L) using the Plackett-Berman sensitivity analysis tool. The PB analyses indicated that total application rate, soil bulk density (an indicator of available water holding capacity), the soil partition coefficient, and pesticide degradation rates were commonly the most sensitive inputs (Table 7-9).

Table 7-9. Results of Plackett-Berman analyses for leaching expressed as relative importance of sensitive components.

	NC1L		NC2L			NC3L		
	Total	Dissolved	Total	Dissolved	Flux	Total	Dissolved	Flux
relative importance of sensitive components								
Bulk density (AWHC)	22	16	51	77	15	20	14	13
Total applied	26	23				20	19	12
Max rooting depth	7	6	3	1	6	8	8	16
Decay rate (layer 1)	20	16	34			7	7	6
Decay rate (layer 2)						4	4	11
Decay rate (layer 3)								2
Kd (layer 1)	4	17		4	60	5	5	
Kd (layer 2)						3	9	6
RO curve number 1			4	2	5			4
RO curve number 2	3	2	5	4	10			
Pan factor		2	4		2	4	3	8

Monte-Carlo simulations were also performed with GA1L and NC4L to evaluate the effect of uncertainty in the input parameters on the model predictions (Appendix 8). For GA1L the predicted soil concentrations seemed to be in good agreement with the measured values. The estimated spatial and temporal profile of pore water bromide movement through the soil core was highly correlated to the measured data from the soil suction lysimeters but the model predictions of the magnitude of the bromide pore-water concentration were not accurate. Similar results were obtained for NC4L except that the predicted soil concentrations were not as close to the measured values.

Conclusions from the Leaching Simulations

- (1) Based on the results obtained from bromide simulations, PRZM 3.12 is capable of simulating reasonable rates of ground water recharge in highly vulnerable agronomic settings when the soil characteristics are appropriately represented. Annual recharge rates of 29 to 59% of precipitation plus irrigation were simulated for the study sites.
- (2) The Rawls and Brakensiek regression equations provided in the PRZM3 manual provided appropriate soil characteristics, at least for the sandy, highly transmissive soils present in all study sites.
- (3) The plant uptake of bromide and pesticide was simulated to be a significant dissipation mechanism for highly mobile chemicals with moderate to long soil half-life values. The uptake of pesticides seems to be higher than would be expected based on the limited data available.
- (4) For foliarly applied compounds, the extent of canopy interception, degradation on the plant canopy and foliar washoff are critical in determining the soil loading. Reasonable estimates of these values for foliarly applied chemicals must be provided for accurate predictions.
- (5) Two of the most sensitive parameters affecting the simulated concentrations of pesticide in soil and soil pore-water were the sorption coefficient and the degradation rate. Appropriate values of these properties must be used in order to obtain reasonable estimates of leaching under field conditions. Moving from cold modeling using "worst-case" physical properties to site-specific modeling using "typical" or "best-estimate" values can result in significantly improving the agreement with actual field data. For NC4L, the agreement for pesticide concentration in soil pore-water changed from 4X for cold simulations to a factor of 2X for site-specific simulations.

- (6) PRZM3 was unable to simulate early detections of pesticide which were attributed to a preferential flow mechanism. Some experimentally observed concentrations were relatively low (less than 1 ug/L) and did not result in repeated detections in shallow ground water. The model was generally unable to simulate this behavior.

Recommendations from the Leaching Simulations

- (1) A standardized operating procedure (SOP) should be developed to guide the appropriate parameterization of PRZM3 for use in regulatory submissions. The SOP should provide guidance on selecting appropriate chemical, soil, agronomic and climatic data to ensure consistent and technically sound modeling results that are acceptable for regulatory purposes.
- (2) Due to known deficiencies in the Thornwaite evapotranspiration routine used in PRZM3, daily evapotranspiration was calculated for each site using a modified Penman equation. The ET routine in PRZM3 should be upgraded to a more reliable algorithm.
- (3) For purposes of model evaluation, hydrodynamic dispersion was set to zero at all depths and fixed compartment sizes were used at various depths in the soil profile as specified in the SOP. This approach created some degree of numerical dispersion which resulted in band broadening of the concentration profile with depth. To improve the simulation of the movement of pesticides, recommendations for appropriate levels of hydrodynamic dispersion should be developed through use of parallel modeling using a Richards equation-based models coupled with the convection-dispersion equation.
- (4) The accuracy of simulating pesticide concentrations at soil depths deeper than one meter could be improved with additional experimental data on the rate of degradation in subsoils. For some pesticides, more sophisticated degradation kinetics may be needed to accurately simulate movement of trace level residues to deeper soil depths.

8.0 Conclusions and Recommendations

The following general conclusions can be drawn from the results of the validation work for use of PRZM3 in a regulatory setting to estimate leaching and runoff.

- The work presented in this report shows that PRZM3 provides a reasonable estimate of chemical runoff at the edge of a field if the uncertainties in the model input parameters are propagated to model output. Uncertainty in the input parameters related to scenario definition (for example, field characteristics and climate) is less important in a regulatory setting than when comparing predicted and observed for determining model performance). Simulations based on the best choices for input parameters (no conservatism built into input parameters) show that simulations are generally within an order of magnitude of measured data with better agreement observed for larger events with better agreement obtained for cumulative values over the study period. When the model input parameters are calibrated to improve the hydrology, the fit between predicted and observed improves (results are usually within a factor of three). The errors obtained from the simulations with the calibrated hydrology is relevant to regulatory applications since a fixed scenario definition implies a fixed hydrology. When conservatism is deliberately introduced into the input pesticide parameters substantial over-prediction of runoff losses occur. Apart from the parameters defining the scenario, the most sensitive parameters affecting runoff losses in the water and sediment phases were partitioning coefficient and degradation rates.
- Simulations with PRZM3 show that reasonable estimates of leaching were obtained in homogeneous soils where preferential flow is not significant. PRZM3 usually did a good job of predicting movement of bromide in soil (soil and soil pore-water concentrations were generally within a factor of two of predicted values). For simulations based on the best choices for input parameters (no built in conservatism), predictions of soil concentrations for pesticides were usually within a factor of three and soil pore-water estimates were within a factor of 11. When the model input parameters were calibrated to improve the simulation of hydrology, predicted pesticide concentrations in soil pore-water were usually within a factor of two of measured concentrations. Because of the sensitivity of leaching to degradation rate, the most accurate predictions were obtained with pesticides with relatively slow degradation rates. When conservative assumptions were used to define input pesticide parameters, predictions of pesticide concentrations were usually at least a factor of two greater than when using the best estimate of input parameters without any built in conservatism.
- The initial work conducted by different contractors showed the importance of having an SOP which defines the selection of **all** model input parameters. What seem like relatively unimportant model parameters such as compartment size can have a major effect on the answers produced. The most satisfactory way to implement regulatory modeling is through the development of a shell which provides all input parameters related to the scenario, with the user providing only the parameters related to the specific pesticide being assessed. This is especially relevant for PRZM, which requires an experienced user to provide input data in card format and to interpret the lengthy tabular output.

The comparison of predicted and observed values indicated a few areas where improvements to PRZM3 could have a beneficial effect. These can be divided into two categories. The first are improvements that help to get better predictions for a given scenario (for example, improved descriptions of processes). The second category consists of improvements that help to better define the scenario being simulated. Improvements in the second category are especially important in more precisely describing the scenario when comparing predicted and observed values.

Process Improvements

Probably the most important area related to runoff modeling for further improvement in PRZM3 is a better description of evapotranspiration. The evapotranspiration routine in PRZM3 typically underestimates actual evapotranspiration resulting in incorrect simulation of runoff and excessive estimates of recharge.

For leaching modeling the most important need is better description of degradation rates in surface and subsoils. Because the amount of predicted leaching for mobile compounds is relatively sensitive to the degradation rate, further work needs to be done to more accurately determine the most appropriate degradation rate to use in a specific situation. Because degradation rates from laboratory studies are often slower than actually occur in the field, use of laboratory data sometimes can result in overly conservative estimates of leaching. A number of potential possibilities could be explored. An example would be to see if degradation estimates could be improved by inverse modeling of field studies to determine temperature and soil moisture corrected degradation rates, which could then be used with an appropriate model accounting for changes in degradation with temperature and moisture. One potential area for long range research might be to better understand biological degradation, including the effects of previous stresses on microbe populations such as cold temperatures and dry conditions. When compounds have the potential to move below the root zone, better procedures for developing degradation rates in subsoils (for both chemical and biological degradation pathways) are also important.

One other process that seems to need improvement is the crop uptake factor. Although plant uptake is usually not measured in field studies, the limited field data suggest significantly lower uptake than predicted by PRZM. Predictions from other environmental models (PELMO, PEARL, and MACRO) are also lower than predictions from PRZM3. Usually uptake will have a relatively minor effect on the answer so another approach would be to not consider losses due to this mechanism.

Separating the effects of degradation and sorption is relatively difficult when comparing predicted and observed concentrations since for a given data point in time and space, similar concentrations can sometimes be predicted using different sets of sorption and degradation parameters. In general PRZM seems to overestimate movement with a portion of the residues staying nearer the surface longer than would be expected based on modeling. Sorption routines that consider increasing sorption with time may be one approach to obtaining better agreement between predicted and measured soil concentration profiles.

A number of different processes that are important under environmental conditions are not accounted for in PRZM. These include various types of preferential flow, transport due to freezing and thawing of soil, and crusting of the soil processes. PRZM3 should not be expected to give accurate predictions when any of these processes are a significant pathway for transport. If predictions under such conditions are necessary for regulatory modeling, additional work will be necessary to best define how such processes can be described mathematically and then these descriptions could be incorporated into future versions of PRZM.

Improvements in Scenario Description

In order to be able to use daily rainfall, PRZM3 assigns a rainfall distribution based on a regional average to each rainfall event. Therefore, this average regional distribution is not likely to represent the distribution that occurred during the actual rainfall event at a study site. Using the actual storm intensity in the model predictions would improve the agreement between predicted and observed data. However, such data is not routinely collected at most U.S. weather stations. For regulatory modeling, incorporating actual rainfall intensity rather than using an average rainfall intensity since the scenario defined by the model does not have to represent an actual situation (only be representative of actual situations).

For leaching proper definition of soil hydraulic properties is essential. The work of the task force indicated that use of hydraulic property measurements from disturbed cores did not provide acceptable results. Correlations based on the Rawls and Brackensiek equation greatly improved predictions although data from undisturbed soil cores should be used if available.

Comparisons of predicted and observed concentrations would also probably be improved by including seasonal variations of the crop cover and Manning's roughness equations, allowing the SCS curve number to vary throughout the cropping season, and permitting the extraction depth to vary by site or over time (all of these items are discussed more fully in the summary of the runoff simulations). These changes, unlike the other potential changes in this section, will help to provide a more realistic scenario definitions, even for modeling of a site that is not necessarily intended to be description of an existing site.